Three empirical approaches to the economics of education

Dissertation

submitted to the Faculty of Business, Economics and Informatics

of the University of Zurich

to obtain the degree of

Doktorin der Wirtschaftswissenschaften, Dr. oec.

(corresponds to Doctor of Philosophy, PhD)

presented by

Katherine Marie Caves

from the United States of America

approved in July 2015 at the request of

Prof. Dr. Dr.h.c. Andrea Schenker-Wicki

Prof. Dr. Uschi Backes-Gellner

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 15.07.2015

Chairman of the Doctoral Board: Prof. Dr. Josef Zweimüller

Acknowledgements

A huge project like a dissertation is never done alone, and I have many people to thank for their contributions. First, I would like to express my gratitude to my advisor Prof. Dr. Dr.h.c. Andrea Schenker-Wicki. Thank you for accepting my request to come here to the University of Zurich and for believing in my capacity as a researcher. Thank you for supporting this work, for representing me, for encouraging me, and for finding the time to address any concern with which I came to you. I would not be in the position to do this work without you.

I am also grateful to Prof. Dr. Uschi Backes-Gellner. Thank you for inviting me into your seminar—the discussions we had there and the things I learned there have honed every skill I have as a researcher. Thank you for your feedback and your guidance. Thank you for finding the time for me even though I am not one of your students. Finally, thank you for being part of my dissertation committee despite your very busy schedule.

I am grateful to my colleagues in the Chair of Performance Management. Thank you Mathias Beck for being my guide through most of my doctorate and for always inviting me to 4pm coffee. Thank you Lukas Schoenenberger for your morning company and your advice as I finish this process. Thank you Adina Bialas for your friendship, German lessons, and my sanity. Thank you Stefano Nigsch for always answering my questions and for your consistent geniality. I am also grateful to other members of the Performance Management/Executive MBA working group, including Dr. Xinhua Wittmann, Dr. Christian Rosser, Dr. Matthias Schaub, Dominique Vionnet, Maria Franco, Thomas Hermida, Danielle Tanner, Laura Gallati, Sabrina Pfizer, and Priscilla Tadres.

I owe many thanks to the other scholars and doctoral students with whom I have written and discussed the research detailed here. Thank you to Simone Balestra, Johannes Meuer, and Christian Rupietta for writing with me—both projects were genuinely enjoyable experiences, which is not always the case. I appreciate the feedback of Dr. Bandara Bandaranayake during the paper on which we worked. I am very grateful to the members of Dr. Backes-Gellner's chair for letting me be part of their group, especially Miriam Rinawi. Many thanks to Jamie Lee Gloor for her friendship and sense of rigor, and to the other members of the HRM chair for being so welcoming. Finally, I am deeply grateful to my mentors, Prof. Dr. Daniella Ann Cook—without whom I would not even be doing a PhD—and Prof. Dr. Ursula Reynold.

Finally, all my love and thanks go to my family and to Martin. Mom, Dad, Elise, Sarah, and Oma, you have all kept me happy, grounded, and above the waterline throughout this adventure. Thank you for your love, support, and Skype calls. Sabine, Herbert, Manuel, and Bernhard, thank you for being my family over here. Martin, thank you for your love, for finding my research as cool as I do, and for keeping me excited about doing good work.

Contents

List of Figures

List of Tables

Glossary

ACARA Australian Curriculum, Assessment and Reporting Authority.

DEECD Department of Education and Early Childhood Development.

fsQCA fuzzy-set Qualitative Comparative Analysis.

HSEE High School Exit Exam.

KNN k Nearest Neighbors.

NAEP National Assessment of Educational Progress.

NAPLAN National Assessment of Progress—Literacy and Numeracy.

NCES National Center for Educational Statistics.

OECD Organization for Economic Cooperation and Development.

QCA Qualitative Comparative Analysis.

Chapter 1

Introduction

Research on education is vital for ensuring that all students have access to the best possible start in life. Research on education that comes from the perspective of economics is also key because of its emphasis on data-driven findings and its applicability for educational policy making. This does not take away from the important theory-building work done by education scholars, but rather attempts to translate that theory into concrete findings that governments, school administrators, and parents can use to inform their choices. In this dissertation, I explore three policy-oriented empirical studies that use educational data available to policy makers.

Nearly every developed country collects data on the progress and performance of its students and schools, and international actors like the OECD (OECD, 2013) regularly assess educational performance both within and across countries. With legislation such as the United States' No Child Left Behind Act, educational data is increasingly granular, standardized, and relevant to policy decisions. I explore three potential means by which educational data can be used to guide educational policy making. Each of the three research projects detailed here takes a different perspective in terms of the method used, the purpose of the analysis, and the timeframe to which the results are relevant; all of them contribute to an understanding of how the glut of educational data might be directed towards improving educational outcomes through policy.

1.1 Methodological Approaches

For each study in this dissertation, I use a different method to assess the issue at hand. The first study, entitled "Is funding enough? A configurational analysis of conditions for school achievement in Victoria, Australia," uses Qualitative Comparative Analysis (QCA), which uses set theory and looks for relationships of necessity and sufficiency. The second, "The impact of high school exit exams on graduation rates and achievement," uses a regression-based interrupted time series analysis in both parametric and non-parametric forms. Finally, "Predicting school achievement reactions to increased funding" uses k nearest neighbors (KNN), a non-causal data mining technique, to explore educational data in a way that can be implemented in real educational policy situations. All three methods derive from different epistemologies and use different strategies to draw meaningful conclusions from the data.

In Chapter 2 of this dissertation, I use QCA to evaluate whether an equity-oriented formula funding scheme is successfully accounting for sources of inequity. Previous work already shows that the funding scheme is reducing the correlation between key conditions and school performance (Miller and Voon, 2011), and I find that no condition or configuration of conditions is necessary for schools to succeed or fail. However, I find one configuration of school characteristics that is sufficient for success and four sufficient for failure. Critically, all of the configurations for failure include the school having a low level of community educational advantage while the one configuration for success requires schools to score highly on this measure. The equity problem may be smaller by correlational analysis, but QCA turns up more nuanced relationships.

QCA was developed by political scientist Charles Ragin in the late 1908s and refined throughout the ensuing three decades (Ragin, 1987, 2000, 2008). Ragin's initial insight was that relationships among conditions in the social world are better expressed in terms of necessity and sufficiency than terms of push-pull reactions; they are set-theoretic rather than correlational. Using this perspective, questions like "which configurations of conditions are adequate to ensure that a school succeeds?" (sufficiency) and "is there any configuration of conditions that a school must match in order to fail?" (necessity).

In addition, QCA is designed to further reflect the realities of the social world by accounting for complex causality in the forms of asymmetry, conjunctural causation, and equifinality (Ragin, 2000). Asymmetry refers to situations in which an incremental positive change in a causal factor does not generate a positive change in the response that is equal to the negative change that would result from a negative change in the causal factor. The implication of this is that the causal conditions for schools to fail are not necessarily the opposite of those that cause schools to succeed. Conjunctural causation occurs when the cause for some outcome is a configuration of conditions that cannot always be decomposed to the effects of each individual condition—an outcome could require conditions A and B together but not occur with either A or B. Finally, equifinality captures the idea that there may be multiple pathways to the same outcome—just as A and B together could lead to some outcome, an independent condition C could do the same. QCA is distinct from correlational analyses and uncovers insights that might remain hidden from more traditional methods of analysis.

The study in Chapter 3 uses an interrupted time series design to assess the effects of high school exit exams (HSEEs) on graduation rates and achievement. We find a positive overall effect of HSEE introduction for graduation rate trends, which is heterogeneous over time. In the year of introduction and briefly after, HSEEs have a negative impact on graduation rates, which is short-lived and becomes positive over the long term. We perform robustness checks using states that do not have HSEEs as a control group. We also estimate a pre-intervention negative effect, suggesting that high schools prepare for the HSEE before introduction. We find no effects for achievement, possibly because of lacking meaningful cross-state achievement data.

American states implemented HSEEs at different times, mostly between 1990 and 2012 (Reardon et al., 2010). The use of an interrupted time series approach allows us to center the timeline of all states on the year when they implemented their HSEE policy and, controlling for state- and year-specific effects, evaluate the effect of the policy over time (Dee and Jacob, 2011). This method is an elaboration of regression analyses and, as such, can be interpreted causally in the strictest and most traditional interpretation of that term. Unlike QCA, which demands that its users learn and accept a set-theoretic conception of causality—one that is equally valid but less widely accepted—this method provides clear and actionable results to even the most skeptical policy maker. Furthermore, the interrupted time series approach allows us to generate new insights of the effects of HSEEs on achievement and graduation rates that had remained hidden in single-state or short-term analyses (Bloom et al., 2001).

The final study, in Chapter 4 of this dissertation, explores the link between educational expenditure and student achievement, which is unproved even though economic theory predicts that increased resources should improve educational quality (Hanushek et al., 2011). Using KNN to predict school performance under different financial conditions, I find that the schools with the highest and lowest predicted scores have very small predicted achievement changes when hypothetically given more funding, while the predicted scores of schools in the middle can be moved by adding resources—movements that are greatly varied and often negative. The general trend in prediction change with additional funding is slightly positive for schools at the bottom and slopes downward until it is slightly negative for the best schools. Schools in the third quartile are the most likely to get worse with more funding.

KNN is an approach from the rising fields of data mining and big data analytics in education (Silverman and Jones, 1989; Tanner and Toivonen, 2010). It is a model-free prediction technique rather than more common model-based exploration methods. This means that it is less subject to the assumptions of researchers and the restrictions of parametric measurement, but also that it is non-causal. KNN finds the most similar schools to a given unknown-scoring school in its set of training observations. Based on the scores of those schools, it predicts the score of the given school. I generate predictions for all of the schools in the data and compare them to how those schools are predicted to perform if they had additional funding. By examining the pattern of changes in predicted achievement level along the achievement distribution, I am able to draw insights about how changes in educational funding relate to changes in achievement for different schools.

All three studies explore similar types of data collected by governments and educational authorities, but do so using different methods and by taking different perspectives on the nature of causality and how insights can be drawn from raw information. The first study uses QCA to examine necessary and sufficient conditions in the complex social world. The second turns a rigorous causal eye to a long-term set of diverse cases in order to identify underlying relationships between policy and educational outcomes. The third study takes an exploratory approach and finds evidence that might help illuminate one of the most troubling mysteries in the economics of education. Furthermore, that study operates in a way that can be used by education policy makers in their work. Methodologically, all three studies provide unique perspectives that reveal novel patterns and insights.

1.2 Substantive Issues

In addition to their different methodologies, each study in this dissertation tackles a different problem in educational research. The first study deals with issues of educational equity and how educational authorities can conceive of and manage groups that seem to be at higher risk. The second study—on HSEEs—examines an existing policy across contexts to uncover its real effects on educational outcomes. Finally, the third study examines educational production and the relationship between funding and achievement; a question that has been at the center of education economics since its inception. The different purposes of each study highlight the different approaches that can be taken to asking questions of educational data.

In "Is funding enough? A configurational analysis of conditions for school achievement in Victoria, Australia," the research question is about equity. Specifically, the project addresses whether there are any necessary or sufficient conditions or configurations of conditions for schools in Victoria to succeed or fail. The existing research on those schools has documented a decreasing correlation between some school background characteristics and achievement as the funding formula seeks to correct for those characteristics (Miller and Voon, 2011), but the question of whether there is a set theoretic relationship remains. Equity—along with efficiency—is one of the key issues of education economics and one with which nearly every school system contends (Hanushek et al., 2011). By examining the problem through the lens of necessity and sufficiency, I am able to avoid the argument that equity-oriented policies are dangerous to privileged high achievers because they might decrease the correlation between privilege and achievement by pushing down the top rather than pushing up the bottom. Instead, this analysis simply looks at what it takes for a school to succeed or fail and allows the policy maker to intervene if, for example, a school needs to be privileged to succeed. QCA's allowance for causal asymmetry means that a policy aimed at reducing the necessity of privilege for success comes at no potential inherent cost to already-high achievers (Ragin, 2000). Equity is a key question in education research, and this project provides a positive means of addressing the issue.

The study on HSEEs is a study of how an existing policy affects key educational outcomes. It is unique among research on the issue because it is able to look at HSEEs across state contexts and time while still drawing meaningful and causal conclusions about the policy in question (Dee and Jacob, 2006). This kind of analysis of how policies affect outcomes in the real world is critical for educational policy makers as they work to refine their education systems. Policy analysis is a major part of the body of education economics research, not least because the family of research methods used by education economists is well suited to disentangling cause and effect relationships. This approach to educational data—which seeks the result of a specific causal actor—is key for helping policy makers and researchers understand the results of specific choices and decisions.

Finally, "Predicting school achievement reactions to increased funding" tackles the issue of educational funding and educational production—the second of education economics' key issues. We know from economic theory that increased resources should mean improved quality, but actual identification of that relationship has eluded the education economics research establishment thus far (Hanushek et al., 2011; Hanushek, 2006, 1986). This study finds a pattern in schools' predicted reactions to increased funding that might explain why this pattern is so difficult to identify: the worst- and best-performing schools hardly react at all to increased funding, while those in the middle have much more volatile predicted scores that can even react negatively to an influx of resources. The paper explores potential mechanisms behind this pattern, including inadequate infrastructure in low-performing schools, diminishing marginal returns in high-performing schools, and potential reasons why the middle schools are just as likely to get worse as better with more funding. This kind of analysis may not be causal, but it does shed light on relationships between resources and output in a way that can inform both research and policy.

Educational data does not serve only one purpose, even within the realms of policy making or theory building. The three major mandates of education economics—equity, policy analysis, and efficiency—can all be addressed by similar data sources. By asking the right questions, both education economists and policy makers can derive insights from data that can be used to address questions of theory and issues of real-world policy.

1.3 Perspectives in Time

Beyond their different methods and purposes, each of these studies addresses its educational data from its own chronological viewpoint. The first study, which uses QCA to address educational equity from a set-theoretic perspective, looks at an educational environment in the present tense to discover what is happening and what can change moving forward. The second study on HSEEs looks across the past to determine the short- and long-term effects of a policy that has been in place for decades in some cases. In the KNN study of educational resources and achievement, the analysis uses hypothetical data to see how school achievement might change in the future if given more money per student. Educational data is not limited in its application to one specific time and can be used to assess the current situation, the performance of a past policy, or the potential effects of a change in the future.

By examining an educational policy question in the present time, policy makers can both assess the effectiveness of past policies and plan for the future. The first study uses a cross section of data from one year to determine how schools are now, which generates insights for how they should be changed. This particular study finds that school achievement is possibly more equitable than it was in the past, but that some school characteristics remain important for performance. A study like this could be performed during one school year while making plans for the next. One role of educational data is to help policy makers make decisions in the short term, and analyses like this are key to informing those decisions.

Education policy must be shaped not only by short-term situations but also by knowledge of the long-term consequences of policy choices. In "The impact of high school exit exams on graduation rates and achievement," we construct a dataset of state-level HSEE policy, achievement, graduation rates, and control data that spans over 20 years. This data is all collected by the government and is readily available to policy makers and education researchers. By looking at the effects of a policy like HSEEs across a long period in the past, we are able to find not only its short-term results but also the trends it creates over the long term. Without long-term studies like this one that look deeply into the past, policy makers might lose the systems they are creating in the individual policies of specific problems. Just as current data is important to adjusting policy in the present, past analyses give perspective on what can improve educational outcomes in a lasting way.

Finally, education policy must look towards what might happen in the future. By exploring the potential reactions of schools to policy decisions in a manner similar to the third study detailed here, policy makers can have an informed idea of what outcomes specific policies might generate. The future is obviously difficult to predict, but the fields of data mining and big data analytics have generated myriad prediction tools and models that are increasingly accurate and reliable. These tools can even be more accessible to the average decision maker than the linear algebra of regression modeling techniques, which means that they could be more widely and appropriately used. Predicting responses to policy in a systematic way is a valuable tool for improving education.

Educational policy is not only about the present, but draws also on what has been learned from the past and what can be seen of the future. These three analyses demonstrate ways that educational policy makers and researchers can assess what is happening now, what happened in the past, and what might happen in the future. All of this contributes to the understanding of how the vast amount of data collected by school systems, governments, and international organizations can be put to use for the purpose of improving educational equity, policy, and efficiency.

Chapter 2

Is Funding Enough? A Configurational Analysis of Conditions for School Achievement in Victoria, Australia

2.1 Introduction

In this chapter[1], we take a configurational perspective to examining school resources and achievement, and use Qualitative Comparative Analysis (QCA) to search for necessary and sufficient configurations of conditions for school achievement. Managing the investment of governments in education is an important issue on a global level, both because these investments represent such significant sums of money and because of the proven importance of education for economic growth (Soguel and Jaccard, 2008). Education develops human capital, which translates into both individual and public returns. In

[1]This chapter is adapted from a paper of the same title, written with Bandara Bandaranayake, Department of Education and Early Childhood Development, Victoria, Australia, reachable by email at bandaranayake.bandara.b@edumail.vic.gov.au or telephone at 03 9637 3669; and Andrea Schenker-Wicki, University of Zurich, Department of Business Administration, Chair of Performance Management, reachable by email at andrea.schenker@business.uzh.ch or phone at +41 44 634 29 10. The paper has been accepted for presentation at the American Educational Research Association Annual Meeting in April of 2015 and has been submitted to *Educational Researcher*

Australia, individual returns to each year of education fall between 4.5 percent and 12 percent (Miller et al., 1995), and increase for higher educational qualifications (Preston, 1997). At the national level, if the average education level of all working Australians were increased by one year, the country's GDP would increase by 8 percent (Doric, 2003). However, the relationship between educational inputs and production is not clear-cut (Hanushek, 2006). School efficiency and equity are critical for maximizing the output of educational institutions, but are difficult to manage.

Currently, the strongest predictor of school achievement is the level of educational advantage of the school's community (Miller and Voon, 2011), which presents concerns for the equity of school resource distribution and its efficiency at improving student outcomes. This study expands on existing research by taking a configurational approach to examining conditions and configurations thereof that are sufficient for school-level educational achievement in Victoria, Australia in order to determine whether efforts to correct for sources of inequity are working.

The state of Victoria explicitly prioritizes equity in its statements and policies. Beyond merely stating a goal for students to achieve regardless of family background, the formula used to allocate funding to Victorian schools includes additional funding for students and schools with greater need based on their location, level of community education background, and size (DEECD, 2011, 2012). If these efforts are working, then heterogeneity in school achievement should not rely on background characteristics included in the formula like the school's community level of educational advantage, size, or the urbanicity of its location, but on the consequences of policy choices made by the autonomous governors of each school—including the number of students per teacher, attendance rate, and fraction of funding acquired through fees, donations, and investments. This paper uses fuzzy-set QCA (Ragin, 1987) to determine whether the funding formula is able to allocate resources to schools in such a way that community educational advantage, urban location, and school size are no longer necessary or sufficient for schools to succeed or fail.

We investigate this proposition and find that no single condition or combination of conditions is necessary for schools to succeed or fail, but there are combinations of conditions sufficient for both success and failure. While school size appears to be accounted for in the formula, the urbanicity of a school's location and especially its level of socio-educational advantage remain important for both success and failure.

2.2 Theory

The field of education economics seeks to define an education production function. The objective of this is to generate a road map for improving educational equity and efficiency that can be implemented by policymakers and administrators. There is debate over the proper elements and functional form of the function—about which inputs matter for cognitive development or test scores and how they interact—but the most basic form of the function reflects its combination of educational factors in a general form.

$$T = f(Student,\ School,\ Other)$$

In the simplified function above, T represents test score performance while *Student, School,* and *Other* each represent vectors of characteristics or inputs. Student characteristics include elements like ability and family background. School characteristics can include school resources like funding and qualified teachers, policies like class size and instruction time, and background factors like the school culture. Finally, other factors capture anything else that might be relevant to the production of cognitive skill such as general geographical location or exogenous shocks like natural disasters (Hanushek, 1986; Hanushek et al., 2011; Todd and Wolpin, 2003).

Traditionally, the approach to identifying this function is to test the individual and interactive correlational effects of variables using linear or nonlinear regression methods. Some inputs—especially intrinsic ability—can be extremely difficult to measure and often rely on identifying assumptions. This is complicated by the theoretical understanding that learning is a process that occurs and builds over time while depending on the environment in which it occurs (Todd and Wolpin, 2003). Further, educational production is an extremely complex process and it is unlikely that there is only one possible route to higher test scores (Cooper, 2005). Even more importantly, there is no reason to assume that educational production is symmetric: that factors that improve outcomes will have proportionally greater effects when increased or will have proportionally negative effects—or any negative effects at all—when decreased (Glaesser and Cooper, 2011). In order to address these concerns and the inherently set-theoretic issues of sufficiency and necessity raised by our research question, we turn to an alternative method for this study.

2.3 Context of this Study

The most prominent and consistent finding in analyses of education production in Australia is that the local level of educational background is the strongest single predictor of school success, accounting for approximately 62 percent of primary school test scores (for more information, see Miller and Voon (2011)). This level of educational advantage is captured by the ICSEA score, which is calculated by the central education authority in Australia. ICSEA values represent a school's local average educational background excluding the wealth of either the students' families or the school. It is defined at the school level, with a national median of 1000 and standard deviation of 100. The formula for calculating ICSEA includes the socio-educational advantage (SEA) of students' families, the remoteness of the school, the percent of indigenous students, and the percent of students who do not speak English as their native language and are also disadvantaged. The dominant element of this formula is SEA, which is calculated using a combination of parent factors like occupation, school- and non-school education, occupation, and language background from enrollment data and local information on similar topics from census data. ICSEA is unrelated to previous performance of the school (ACARA, 2012). This value has extremely strong predictive power for school success.

Generally, schools in urban areas do well, as do larger schools, though there is some debate on the relative effects of educational inputs on outcomes (Krueger, 2003). Johnson et al. (2004) use school-level Victorian data similar to that in this study and perform multivariate analyses using primarily a random effects model. For primary schools, larger class sizes have a slightly negative impact for both metropolitan and non-metropolitan schools, but the effect is often insignificant. Lamb et al. (2004) find regional differences by state in educational achievement, but also confirm that schools in metropolitan areas generally outperform those in provincial, rural, or remote areas.

The role of funding in the achievement of schools in Australia yields mixed findings. Johnson et al. (2004) differentiate between locally raised funding and global school budget, but are unable to find significant or consistent results across analyses. Similarly, Lamb et al. (2004) examine core funding, locally raised funding, and special funding separately but find either insignificant or inconclusive results. Marks (2010) decides that resources are irrelevant to student performance. This is consistent with international findings which, taken together, show no clear trend of educational inputs directly affecting outputs or outcomes (Hanushek et al., 2011). Despite emerging evidence supporting the importance of financial resources for school success (Dewey et al., 2000) and possibly because of the way schools are funded in Victoria, the role of different funding sources in Victorian educational achievement is still an open question. These individual effects provide an opportunity to assess the configurational effects of educational inputs interacting.

Schools in Victoria are funded using a formula called the Student Resource Package (SRP) that accounts for a basic per-student allocation, school-maintenance allocations, and special funding initiatives for students with greater need (DEECD, 2011, 2012; Bandaranayake, 2013). Funding by formula has the advantages of adding transparency and accountability to educational funding, and works best when schools have autonomy in allocating the resources given to them (Soguel and Jaccard, 2008; Levacic, 2007; Schenker-Wicki, 2008; Levacic and Downes, 2004; Levacic et al., 2000). Victorian schools are free to collect, invest, and spend money as is determined best by their principal and school council. The vast majority of government funding is state-based, at 80-90 percent. Non-government funding is reported as either fees and parental contributions or other private sources like donations and returns from investments.

The purpose of a funding formula is to reflect the market price of educating a given student to a certain level (Levacic, 2007). Victoria's SRP is made up of core student funding, equity funding, and allocations for school infrastructure, school-specific programs, and targeted initiatives. The core student learning allocation delivers a certain dollar amount per student based on their grade level and some school characteristics including size and the remoteness of their locations. Equity funding provides additional funding for students with disabilities or special needs and for those with non-English language backgrounds and low levels of family educational advantage—this is the element of the formula that attempts to correct for the variables reflected in the ICSEA score. School infrastructure funding reflects the maintenance, cleaning, and energy costs of the school. School-specific programs include special programs like music or art classes and other infrastructure needs specific to individual schools. Finally, targeted initiatives are short-term or directed grants used for experimental programs or further equity concerns (DEECD, 2012).

Taken together, these reflect the explicit orientation of schooling in Victoria towards equity. School location and size are both explicitly accounted for, as are the variables that compose ICSEA score. This leads us to our research questions: Does the funding formula in Victoria adequately account for school size, location, and ICSEA score such that these are neither necessary nor sufficient for school success or failure on the NAPLAN assessment? If so, are these relationships affected by teacher-student ratio, attendance rate, or the collection of non-government funding?

If the formula accurately reflects the price of educating each student to the same standard, these conditions should not be necessary or sufficient for schools to do well or poorly on exams. We expect that membership in the sets of high attendance rate, high teacher-student ratio, and high ratio of local to government funding will all be part of configurations related to school success, while the inverse will be true for failure.

2.4 Empirical Strategy

Qualitative Comparative Analysis QCA is a set-theoretic tool that uses Boolean algebra to perform comparative analysis of cases (Ragin, 1987, 2000, 2008). The analysis uses a truth table of all possible configurations of conditions and their associated consistencies with our goal outcome to determine which conditions or configurations of conditions are necessary or sufficient for school success, measured in this case by higher test scores. Policymakers use a language of necessity and sufficiency, so research methods should actively reflect that and the concerns implicit in it. In cases when theories are formulated in terms of necessity and sufficiency, QCA is the best option to answer the questions that arise (Schneider and Wagemann, 2012). We will provide a brief overview, but we refer the reader to Ragin's publications for more thorough explanation.

In set theory, sufficiency and necessity are subset relations. A is sufficient for B when A is a subset of B: not all instances of B include A, but all instances of A include B. A is quasi-sufficient for B when A is almost entirely a subset of B. Conversely, A is necessary for B when B is a subset of A, and quasi-necessary for B when B is almost a subset of A. Set language is frequently used in research on education, with factors described as necessary or sufficient. By explicitly using these ideas, we are able to answer the question of whether Victorian funding compensates for community educational background and location or it is still necessary or sufficient for school success. For a more detailed explanation of the subsethood relations implied in sufficiency and necessity, see Glaesser and Cooper (2011).

QCA has some advantages in the context of education research. First, the method does not assume that conditions' effects are symmetric: the presence of a certain condition may be sufficient for an outcome, but its absence is not necessarily sufficient for the negative outcome. There is no reason to make this assumption in education, and it allows us to capture the differences in issues that struggling schools are facing from those of more successful schools. In addition, the configurational nature of QCA allows us to examine conditions as they function together rather than individually. For conditions with unclear empirical results like school size and teacher-student ratio, this will be helpful in discerning mechanisms through which they affect school performance in conjunction with other conditions. Finally, QCA allows for multiple pathways to a given outcome, so the configurations returned by this method of analysis are an immediately applicable road map for school authorities, administrators, and leaders to follow. A school can identify the configuration that matches them most closely and work to meet—or avoid, in the case of configurations sufficient for the negative outcome—the specific benchmarks of the configuration.

QCA works by laying out every possible configuration in a truth table and populating that with data. Using Boolean algebra, it identifies factors that are consistently present or absent between configurations with the same outcome, disregarding those that differ. The logic of this is that factors whose presence and absence both lead to the desired outcome must be irrelevant for that outcome. By reducing all possible configurations to only those factors that must be present or absent for the outcome in question, QCA comes defines one or more configurations—combinations of factors and/or the absence of factors—consistently sufficient for the outcome.

QCA originated in political science, but it has been used to examine efficiency and equity in education. Researchers have used the method to assess the effects of student characteristics, teacher characteristics, and educational policies on student outcomes (Cooper, 2005; Cooper and Glaesser, 2008). The method has been useful for analysis of equity in educational success in the United Kingdom (Cooper and Glaesser, 2010; Cooper, 2005; Glaesser and Cooper, 2013) and of policies on track switching in Germany (Cooper and Glaesser, 2011). Poder et al. (2013) use fsQCA to find three causal pathways to maximize efficiency and one for equity in European education systems. Recently, Trujillo and Woulfin (2014) used the method in conjunction with qualitative work on school reform intermediaries in California.

2.5 Data

Data was collected for all schools in the state of Victoria, Australia through the Australian Curriculum, Assessment and Reporting Authority (ACARA). ACARA collects and maintains data on school test scores, finances, background, and policies. We limited the sample to government-operated primary schools with complete data—only 26 schools were eliminated for incomplete data. We chose government schools because independent and catholic schools are subject to different funding and reporting laws and we do not wish to bias our results. Furthermore, there is evidence that a selection bias generates a greater share of test score differences between government and non-government schools than school performance (Miller and Voon, 2011). Primary schools make up the vast majority of government schools in Victoria (over 77 percent of government schools in the sample), and test scores can be compared without bias—test scores in 9th grade likely require something different from schools than those in 3rd and 5th grades. The final sample includes 1015 schools.

Nationally, ICSEA scores range between 500 for extreme disadvantage to 1300 for extreme advantage. ICSEA values in our sample ranged between 740 and 1209 points, with a mean of 1017 and a median of 1000. As mentioned earlier, ICSEA value has been identified by previous research as the single strongest predictor of school success in this context. From a theoretical perspective, this makes sense: the educational background of a school's community is related to student family background and school factors, so improvements in ICSEA score affect school achievement from multiple perspectives. The dialogue of educational achievement in Australia implies that having a high ICSEA value would be sufficient for a school to be successful, and that such a score may be necessary for success (Miller and Voon, 2011).

Urban location is another strong predictor of school-level outcome in Australian education research (Marks, 2010), but its effects are often attributed to the generally higher socioeconomic status of urban communities in that context. Theoretically, a metropolitan school would have greater access to resources and would be able to add these to their educational production and boost performance. Given the claim that urbanicity is useful because of the access to resources that it implies, we predict that urban location will be insufficient by itself for school success but will be sufficient for success when combined with other resource-related conditions.

School size is more difficult to fit into the model of education production because the size of the school does not inherently affect the level of resources distributed to each student. In line with this, findings on school size have been weak (Miller and Voon, 2011). We include school size in the analysis because it may combine with other resources—for example a high teacher-student ratio—in successful configurations without being sufficient for success or failure on its own. This is an area where configurational analysis can be particularly useful—school size by itself is not immediately relevant for school performance, but it may be important in the presence or absence of other conditions.

Teacher-student ratio—a proxy for class size—is a topic of debate in the economics of education. Earlier models of education production assumed that students in smaller classes would get a greater fraction of the teacher's attention and use that resource to generate greater performance. However, research on class sizes cannot find consistent evidence that smaller class sizes directly affect student performance in the short term or over time (Krueger, 2003; Jackson and Page, 2013). We include the condition because, like school size, a configurational analysis may reveal areas where teacher-student ratio is important for success or failure as part of a larger pattern even though evidence for its individual correlational effects is mixed.

Attendance rate measures the amount of exposure students get to school resources, which in turn affects education production. Most schools in the sample fall within a fairly small range, but there is enough variability that attendance rate may act in conjunction with other factors as part of configurations sufficient for success or failure. Financial data collected by ACARA records each school's income by origin, from both government and non-government sources. As mentioned earlier, the vast majority of funding comes from the government through the SRP. Government schools generally collect less than 2,000 AUD in total from fees, charges, and parent contributions; other private sources—including the school's investments—are varied but frequently account for more than 100,000 AUD per school per year. We summed federal and state funding to generate total government funding, and the two non-government categories were combined into local funding. We divided these by the total number of full time-equivalent enrollments at the school, yielding per-student government- and local funding. The condition used in analysis is the ratio of local to government spending per student

The ratio of local to government funding does not capture absolute resources, but it does capture extra financial resources from the community after correcting—through the funding formula—for the school's level of need and the cost of educating its unique student population to the level expected by the Australian government (Levacic, 2007), measured by scores on the NAPLAN test. Thus, the condition of local funding per government dollar (LPGFUND) measures additional resources beyond those deemed necessary for a school's unique needs and can be interpreted as a measure of resources beyond the basic necessities that controls for school-level heterogeneity of need. While the importance of financial resources for school outcomes is predicted in theory, evidence for the correlational effects of financial resources on educational output is inconclusive and mixed (Hanushek et al., 2011). The success of funding in generating results in any system including one that uses formula funding is reliant on the quality of the system's implementation (Schenker-Wicki, 2008; Levacic, 2007), so we predict that higher local funding per government dollar will be present in configurations sufficient for success, but that it will not appear alone.

A standardized exam is used throughout Australia, called the Australian National Assessment Program—Literacy and Numeracy (NAPLAN). NAPLAN was introduced in 2008 as an annual assessment for all students in Years 3, 5, 7 and 9 on skills in reading, writing, spelling, grammar and punctuation, and numeracy. This is the primary means by which the central educational authorities in each state maintain school accountability. Our data on NAPLAN scores is at the school level by grade (3 and 5) and subject (literacy and numeracy); we combined these into a single average score. Variation of standardized exam scores is consistent with the approaches in the literature to measuring performance in primary schools (Johnson et al., 2004). Furthermore, Australian educational research has indicated that results are generally insensitive to the type of NAPLAN score used as the outcome variable (Miller and Voon, 2011). Descriptive statistics are in Table 2.1.

Table 2.1: Descriptive Statistics

N = 1,015	Mean	St. Dev.	Min	Max
ICSEA Value	1,017.178	70.597	740	1,209
Teacher-Student Ratio	0.069	0.017	0.048	0.200
Attendance Rate	93.831	1.547	87	97
Government Funding	9,341.908	2,934.124	3,523.822	31,480.470
Local Funding	767.352	375.825	195.246	3,077.706
Local/Government Funding	0.087	0.044	0.013	0.348
NAPLAN Score	455.471	30.430	347.000	537.500

2.5.1 Calibrations

Calibration is the determination of set membership criteria, and is an important part of fsQCA. Set theory differentiates between crisp and fuzzy sets. Set membership options in crisp sets are limited to 0 (fully out) and 1 (fully in). We use this calibration style for school location. In fuzzy sets, membership scores can cover the entire interval from 0 to 1, where 0 is fully out, below 0.5 is more out than in, 0.5 is the crossover point at which the case is neither in nor out, above 0.5 is more in than out, and 1 is fully in the set. This is useful when variables are continuous, such as test scores or funding dollars. In calibration, we decide where those points fall so that cases' features can be assigned set membership levels.

In this study, set membership was calibrated using the indirect method (Ragin, 2008), wherein three anchor points are established for fully out (0), the crossover point of maximal ambiguity, and fully in the set. For most sets, we used the 10th, 50th, and 90th percentiles. For test scores, we used quartiles (25th, 50th, and 75th percentiles) to reflect how school performance is most often reported in this context. This style of calibration is useful in larger datasets because it is easily replicable and because the size of the data increases variability and eliminates clear discontinuities that could be used for manual calibration. Calibration cutoff points for all sets are shown in Table 2.2. Truth tables for both analyses can be found in the appendix Table A.1 and Table A.2). Data analysis was conducted using the QCA package for R (Thiem and Dusa, 2013).

School background conditions include the school's average level of community edu-

Table 2.2: Calibration Cutoffs

Condition	Set Name	Fully Out	Crossover	Fully In
ICSEA Value	ICSEA	945	1,000	1,127
Location	URBAN	Non-Metro	-	Metropolitan
School Size	SIZE	54.2	249.0	554.6
Teacher-Student Ratio	TEACH	0.0418	0.0772	0.1471
Attendance Rate	ATTEND	0.92	0.94	0.95
Local/Government Funding	LPGFUND	0.04	0.08	0.15
NAPLAN Score	OUTCOME	435.00	454.25	478.63

cation background, location, and size. Community education background—measured by ICSEA value—is calibrated using the 10th, 50th, and 90th percentiles. Location can be metropolitan, provincial, or remote. Victoria is a fairly developed state; so most of the schools are either metropolitan (570) or provincial (444), with only one school classified as remote. This is a categorical rather than continuous variable, so we define the crisp set URBAN congruent with the approach of Johnson et al (2004). A metropolitan school is in with a set membership of one and a provincial or remote school is out with a membership of zero. Finally, school size is also directly calibrated using 10th 50th and 90th percentiles and named SIZE.

Other school characteristics fall more under the control of the school—often not completely, but they are factors upon which the school can act—and include staff-student ratio, attendance rate, and non-government funding. Student-staff ratio is calculated by dividing the full time-equivalent teaching staff by the full time-equivalent enrollment of the school. This set is called STAFF and directly calibrated using the same percentiles as above. The set of attendance rates is called ATTEND and is also directly calibrated. The degree of additional financial resources as discussed above is captured in the set of high local funding per government dollar. This is named LPGFUND—for ?Local Per GOVernment FUNDing?—and directly calibrated such that schools in the 90th percentile and above are fully in, schools above the 50th percentile are more in than out, schools below that threshold are more out than in, and schools below the 10th percentile are fully out.

Finally, the outcome is the school's average NAPLAN test score for literacy and numeracy subjects. These are calibrated using the first quartile, median, and third quartile values to reflect the classification strategy employed by Victorian educational authorities. Data analysis was conducted using the QCA package for R (Thiem and Dusa, 2013).

2.6 Results

Analysis for necessary conditions for both successful and unsuccessful schools found no single condition or combination of conditions that met the standards of consistency required for reporting. There is no condition or combination of conditions of which all successful or unsuccessful schools are a subset. This is good news for the accuracy of the formula for determining the price of educating diverse students and for the equity of the Victorian education system, but is only half of the story.

We move on to assessing which configurations of conditions are sufficient for schools to be successful where the set of high-achieving schools is those scoring at least above the median on the NAPLAN test to be more in the set than out, or in the top quartile to be fully in. To ensure the quality and utility of the results, we established frequency and consistency cutoffs that will be used consistently across both positive- and negative-outcome analyses[2]. The frequency cutoff limits the analysis to truth table rows that represent at least five schools. The consistency cutoff requires that a truth table row be at least 80 percent consistent in leading to a successful outcome in order to be classified as a successful configuration.

[2]QCA can be overly sensitive to individual cases, so common practice is to use frequency cutoffs to ensure that the results are driven by truth table rows containing at least some number of cases. In this case, the combination of frequency and consistency cutoffs means that only about two thirds of the school population enters any given analysis. This could potentially limit the diversity of our results, but we choose to keep the cutoffs in order to provide more consistent and usable results while avoiding the bias of an individual case.

2.6.1 The Positive Outcome

The first analysis, with basic school background conditions, returns one configuration sufficient for higher NAPLAN scores. The results are shown in Table2.3. Often, a solution will be composed of multiple possible configurations, where a successful school can match one or another.

Consistency and coverage are simple to understand in crisp set analyses, but quickly become complex and even paradoxical when using fuzzy sets. In a crisp set analysis, coverage and consistency represent the fraction of successful cases covered by a configuration and fraction of cases matching the configurations that will be in the set of successful schools, respectively. The coverage and consistency of the solution have the same meaning, but for the solution as a whole. Finally, the unique coverage of each configuration represents the proportion of successful schools that match only that configuration and no other configuration in the solution. Schools can match more than one configuration in a solution, so coverages do not need to total to 100 percent (Ragin, 2008). With fuzzy sets, these numbers are not so directly interpretable. A case that is more out than in a set can still have membership of 0.3, for example, and this nonzero membership is included in consistency and coverage scores (for more on this, see Cooper and Glaesser (2011)). We will use caution in interpreting coverages and consistencies, but generally higher is still stronger.

Table 2.3: Solution for High Achievement

Solution	Raw Coverage	Unique Coverage	Consistency
ICSEA*URBAN*LPGFUND	0.495	-	0.900

The combination of high community socio-educational advantage with an urban location and a relatively high level of non-government funding beyond the price of educating students determined by the formula is sufficient for school success. In fact, this configuration is the only consistently sufficient pathway for success, meaning that schools without these three factors do not have a reliable configuration of school characteristics and policies to follow for success. The solution does not cover all successful schools—schools without the configuration can and do succeed—but there is no other reliable pathway.

2.6.2 The Negative Outcome

Just as the features that appear in the configuration are interesting, so are those that are absent. School size is irrelevant, indicating that it is adequately accounted for in the formula. Similarly, teacher-student ratio and attendance rate are irrelevant, which matches the growing prevalence in Australian education research of the finding that within-school administrative and instructional choices are far more relevant for achievement than class size policies and small variations in attendance rates (Jensen and Sonnemann, 2014). This finding does expose a mechanism for funding in which it is sufficient for success when combined with other school factors. By using a configurational method, we were able to find new results.

The configurations sufficient for higher school achievement are asymmetric to those for lower school achievement, so we repeat the analysis with the schools scoring in the bottom quartile as the outcome set. Using the same calibrations, frequency cutoff of five, and consistency cutoff of 0.8, our sample for this analysis was 639 schools. Results for the analysis of sufficient conditions for low achievement are in Table 2.4.

Table 2.4: Solution for Low Achievement

Solution	Raw Coverage	Unique Coverage	Consistency
icsea*attend	0.575	0.119	0.849
icsea*teach*lpgfund	0.417	0.000	0.844
icsea*URBAN*TEACH	0.240	0.004	0.843
icsea*size*LPGFUND	0.297	0.017	0.779
Solution	0.727	-	0.818

Here, we find stronger effects of community socio-educational background and interesting configurational results for other conditions. There are more routes consistent for failure than there are for success—Tolstoy might have been talking about schools when he said that "all happy families are alike; each unhappy family is unhappy in its own way" (Tolstoy, 1877). These four configurations together form a solution with high consistency and coverage: most likely, schools that match at least one of these configurations will struggle, and many struggling schools match at least one of these configurations. One key point here is that low ICSEA score is present in all four configurations.

The first configuration has the highest raw and unique coverage, meaning that it represents the greatest number of failing schools. It combines low ICSEA score with a low attendance rate, and the presence of all other conditions is irrelevant. These may be schools where the community is not highly educated and the students are not getting to school, either because school attendance is not prioritized or because it is difficult. These students have less access to education at home and at school, affecting the background and school aspects of their educational production regardless of whether or not the school is in an urban location, its size and teacher-student ratio, and its local fundraising capacity.

The second configuration combines low ICSEA score with low teacher-student ratio and a low ratio of government to local funding. This configuration has relatively high raw but no unique coverage—this is a possibility when using fuzzy sets and is not alarming (see Ragin (2008) or Cooper and Glaesser (2011) for more information). This configuration represents a general lack of resources: schools with larger classes or fewer teachers, low additional funding, and less-advantaged educational backgrounds. The low ICSEA score implies that students' background factors will be poor, and the school may not have the resources to make up the difference.

The third configuration combines low ICSEA score with an urban location and a high teacher-student ratio. While it seems counterintuitive that urban location and small class sizes would be associated with a struggling school, it is important to remember that findings on urbanicity are strongly tied to the typically higher socioeconomic status of urban communities in Australia, and findings on class sizes are inconclusive at best (Miller and Voon, 2011). One possible interpretation is that these are urban schools in disadvantaged neighborhoods—indicated by the low ICSEA score. When all of the relevant conditions are taken together, a school in this configuration could be an urban school with a disadvantaged community where the smaller class sizes are due to previous attempts at intervention or a greater need for teaching assistants to manage students with additional needs. Either way, the smaller class sizes are failing to help improve outcomes (Jensen and Sonnemann, 2014).

The final configuration combines low ICSEA score, small size, and a high ratio of local to government funding. Because of the formula funding scheme, schools in Victoria have almost complete autonomy within the framework of the law. Specifically, they are free to spend and invest their government-issued funding, and to collect additional funding from fees, parents, and other private sources. A possible interpretation of this configuration is a small school with good financial resources and support that struggles to appropriately invest its resources in policies and practices that promote student achievement on the NAPLAN test.

2.7 Discussion

If anything, the solution for unsuccessful schools offers more insight than that for successful schools. First, while ICSEA is neither sufficient nor necessary for school success on its own, its presence in all four configurations in the failure solution raises equity concerns that were not covered in the success analysis. Despite the greater coverage of the failure solution and the increased diversity of pathways, there is no consistent pathway to failure for schools with high ICSEA scores. Put differently, even though it may be possible for schools with low ICSEA scores to succeed, schools with a high ICSEA score are all but insulated from failure—low-ICSEA schools are not precluded from success, but high-ICSEA schools are nearly protected from failure. Even if the funding formula and education policies in Victoria are reducing the positive correlational effect of community educational background on school success, they do not appear to be making up for its role as a negative actor as effectively. This asymmetric finding is especially relevant for the discussion of equity in education.

In these configurations, neither urban location nor school size are consistently present or absent. Urban location is present in the single configuration for successful schools, but not consistently absent among failing schools. The funding formula may be accounting for urban location better in struggling schools than it does in succeeding schools, possibly because the advantage of an urban location is asymmetric. When an urban location is likely unrelated to its usual higher socioeconomic status, it is present in a failing configuration. With a configurational perspective, these distinctions can be teased apart. Similarly, school size makes its only appearance in a configuration with low ICSEA value and high local funding, indicating that its effects may also be configurational and reliant on combination with other factors. Beyond highlighting the utility of a configurational perspective, these observations demonstrate that the funding formula is accounting for urbanicity and school size effects in many cases—though not always.

The failure analysis also provides more insight into the role of controllable factors than the successful analysis. Low attendance appeared as part of a failing configuration, as expected. Teacher-student ratio appeared in both its presence and absence—again reflecting the lack of evidence for small class sizes as a policy tool (Jensen and Sonnemann, 2014). Additional local funding was similarly mixed, meaning that its effects may also be configurational, asymmetric, and equifinal in nature. Further analyses that take on the perspective of complex causality should prove useful in demystifying its effects.

In order to test the robustness of the model and following the recommendations of Glaesser and Cooper (2013), we tested the modified calibrations, frequency cutoffs, and consistency cutoffs. Our results were impressively reliable. This reliability indicates that the results are not a coincidence of our model specifications but a reliable outcome of the data itself.

2.8 Conclusions

The main finding of this study is that no condition is sufficient or necessary for school success alone, and no factor or combination of factors is necessary for school success or failure. One configuration of school conditions is sufficient for success, and four configurations are sufficient for failure.

The combination of high community socio-educational advantage with an urban location and a relatively high level of non-government funding is sufficient for school success. In fact, this configuration is the only consistently sufficient pathway for success. The solution does not cover all successful schools—schools without the configuration can and do succeed—but there is no other reliable pathway. School size is irrelevant, and funding does appear to play a role in school achievement in specific combination with other school factors. By using a configurational method, we were able to find new and theory-supporting results.

In the four different configurations that are consistently sufficient for low achievement, all four include low ICSEA scores. While low ICSEA is not sufficient for failure by itself, there is no consistent pathway for schools with high scores to fail despite a large majority of unsuccessful schools being covered by the solution as a whole. Schools with low ICSEA are not guaranteed failure, but they cannot match the one configuration that is reliably sufficient for success and must seek out their own alternatives or less-consistent options. This raises equity concerns that may be invisible in other analyses.

Urban location and school size are much less important than socio-educational advantage—not appearing in all but one configuration each—and their presence highlights possible blind spots in the formula based on the configurational and asymmetrical behavior of school factors. Controllable factors like teacher-student ratio, attendance rate, and the amount of funding schools collect are all similarly complex in their causality. By examining these conditions configurationally, we were able to assess possible mechanisms in a new way.

Although our results are illuminating and robust, there are limitations to consider. We measure achievement in terms of exam performance, not life skills or other less measurable outputs of schools. Our data for this initial exploration is limited to the school level and to the conditions available in our data. We cannot analyze outcomes for individual students, nor can we account for the socioeconomic background of the school. Future studies of this nature should address these issues. Additionally, the interpretation of the local to government funding ratio relies on the assumption that the funding formula is—as it aims to be—an effective equalizer. Another possible interpretation of this funding condition is to assume that it represents the economic status of the school's community: schools with low government funding due to their high status and low need are the same schools that collect the largest donations from the deep pockets of their community. This interpretation does not change our conclusions, but it should be taken into account. Given that these are government schools and not the many Catholic or independent schools in Victoria, we are comfortable with our interpretation.

The DEECD in Victoria explicitly sets equity and achievement as its key goals. In the case of the negative outcome especially, there are indications that equity concerns are not adequately met by current interventions. However, there is also evidence that government funding is going where it is needed, as nothing is necessary for success or failure. However, for the lowest-performing schools, additional educational interventions may be necessary to correct for the deleterious effects of low socio-educational advantage.

Chapter 3

The Impact of High School Exit Exams on Graduation Rates and Achievement

3.1 Introduction

As part[1] of the increasing trend towards school accountability and standards-based education over the past two decades, most American states have implemented high school exit exam (HSEE) policies requiring that all high school students pass a test to graduate (Reardon et al., 2010). Exit exams ideally ensure a minimum achievement level for high school graduates and raise the value of high school diplomas, but many scholars worry about the negative effects of such exams on student motivation, graduation rates, and equity (Dee and Jacob, 2006). Estimation of the real effects of high school exit exams is difficult because randomized experimental application of the policy does not exist, so disagreement remains on whether exit exams have positive effects—increasing the achievement level and degree value of high school graduates—or negative effects—

[1] This chapter is adapted from a paper of the same title, written with Simone Balestra, University of Zurich, Department of Business Administration, reachable by email at simone.balestra@business.uzh.ch and by telephone at +41 44 634 42 20. The paper has been accepted for presentation at the American Educational Research Association Annual Meeting in April of 2015 and is under review at *Educational Evaluation and Policy Analysis*

decreasing graduation rates and fostering inequity. Although discussion of exit exams' effects often centers on graduation rates, it remains unclear whether or not requiring an exam negatively affects the graduation prospects of high school students.

In this chapter, we exploit the staggered implementation of HSEEs at the state level over the period from 1990 to 2013 to examine their real long-term effects on graduation rates and achievement and how those effects persist over time. We center each state's time line on the year in which high school diplomas were first withheld based on exam scores. In the linear specification, we find a downward jump in graduation rate in the first year of withholding that recovers completely within five years. In the non-parametric specification, we find a short-term decrease in graduation rates both preceding and immediately following the application of the exam, but graduation losses are recovered within four years and continue improving due to the increasing slope of the graduation rate trend after the HSEE.

Previous findings indicate that the implementation of exam policies increases dropout rates, especially for minority and low-income students (Bishop and Mane, 2001; Dee and Jacob, 2006; Jacob, 2001), although Warren and Edwards (2005) find no effects. The average effects of educational policies including HSEEs can be misleading as they often have distributional effects (Jackson and Page, 2013; Ou, 2010). Dee and Jacob (2006) find that exams reduce the probability of completion overall, but that these effects are particularly strong for black students and in urban school districts with high levels of poverty or minority enrollments. At the same time, they find that HSEEs actually lower dropout rates in more affluent districts.

For graduation as opposed to dropout rates, the finding of no effects is more prevalent (Carnoy and Loeb, 2002; Greene and Winters, 2005; Grodsky et al., 2009; Warren and Edwards, 2005), again with some exceptions (Amrein and Berliner, 2002; Marchant and Paulson, 2005). Establishing causality, however, is difficult and there have been some methodological shortcomings; work focused on causal analysis finds a slight increase in dropout rates among black and low-income students with HSEEs (Dee and Jacob, 2006; Murnane, 2013; Warren et al., 2006). Overall, there is some evidence that HSEEs can improve student achievement and the attainment of high school diplomas, but this effect is highly dependent on school resources and subject to equity issues.

3.2 Theoretical Background

The effects of HSEEs on graduation rates can be approached from two distinct theoretical perspectives: the economics of education and the sociology of education. From an economic theory perspective, increasing the difficulty of acquiring any qualification will decrease the number of individuals achieving that qualification. Importantly, qualifications gain value as a signal of ability when they are difficult to attain (Arcidiacono et al., 2010; Tyler et al., 2000; Spence, 1973). From a sociological and pedagogical point of view, the effects of HSEEs are less clear. Exams could potentially enhance student achievement by focusing school curriculum and teacher instruction on the relevant standards, encouraging the provision of targeted assistance for low achievers, and motivating students (Bishop, 1997; Bishop et al., 2001). Conversely, the exam could demotivate lower-achieving students, incentivize schools and teachers to give up on "hopeless cases," and enhance inequity especially among non-English-speaking students (Booher-Jennings, 2005). If the HSEE is based on flawed standards or fails to adequately measure the standards it seeks to assess, all of these issues are compounded.

Economics predicts positive overall effects of HSEEs for the value of a high school degree and the future prospects of graduates, but acknowledges that these benefits come at a cost of lower attainment. The predicted benefits of HSEEs are non-negligible for both graduates and as a tool for school policy. Achievement of a known minimum standard creates accountability and meaning for a high school education. The increased standards and achievement raise the value of the diploma as a signal of graduates' ability, which improves the labor market and college application prospects of graduates by reducing informational asymmetries in the labor and college market. At the school policy level, the HSEE provides a means of measuring school performance, making schools accountable for teaching the material for which they are responsible (Bishop and Mane, 2001; Reardon et al., 2010). The state can be sure that tested schools will work towards achieving the standards on which they are tested.

However, these benefits are inextricable from lower attainment: an HSEE that prevents the lowest achievers from graduating will increase average achievement among high school graduates by mathematical necessity even if it has no effects on the behavior of students or schools. Increased standards raise performance, but at least part of this effect comes from increased selection. HSEEs are an advisable policy from an economics of education standpoint, but they do prevent the lowest achievers from graduating.

The sociological perspective on HSEEs is more behavioral than incentivist in its predictions for how schools, teachers, and students will react to the implementation of an HSEE and is more skeptical of the accuracy of measurement and appropriateness of standards contained in the exam. As a result, the value of an HSEE and its likely effects on graduation rates are more complex and unpredictable. Positive outcomes of HSEEs revolve around their ability to focus school, teacher, and student efforts (Bishop and Mane, 2001; Bishop et al., 2001). Schools and teachers that know students will be tested on specific material will orient curriculum and instruction around ensuring that students master that content. Students at risk of failing can be given targeted assistance, which can include language services for students whose native language is not English. Students, aware that they will be held responsible for what they learn in class, may be more motivated to study and retain key concepts.

Alternatively, however, HSEEs can generate negative incentives that undermine school, teacher, and student behavior. Schools and teachers may take content specificity too far, yielding curriculum and instruction that covers only what is on the test without substantive context or meaning. Teachers may decide that some students have no hope of passing and prioritize them lower than those on the bubble (Booher-Jennings, 2005). Students who believe they cannot pass may be demotivated or incentivized to drop out. Regardless of the behaviors of schools, teachers, and students, lack of resources in some schools and districts might prevent the implementation of necessary changes, creating a social justice problem (Plunk et al., 2014). There are strong theoretical arguments for both positive and negative effects of HSEEs on school, teacher, and student behavior and performance.

Standards or exams that do not adequately represent or measure appropriate educational goals are an enormous obstacle to the success of HSEEs from any perspective. Educational standards are difficult to set, especially given that the purpose of education is contested in the United States. A successful high school education can be defined as preparing students for any or all of entering the labor market, attending college, acting as informed citizens, and functioning in society. States define curricula for the material and level of mastery required of their high school graduates, and there is no guarantee that these standards or the goals they represent are necessary or sufficient for graduates to succeed in later life. This is further complicated by the potential for HSEEs to fail at accurately measuring the achievement of those standards. Schools, teachers, and students are incentivized to pass the exam, but the skills necessary for that goal may not match the curriculum standards or relevant skills if the HSEE itself is poorly conceived (Akerlof, 1970; Gibbons and Katz, 1991).

HSEEs focus school, teacher, and student energy on mastery of whatever skills and knowledge are required to pass. Exams can incentivize schools and teachers to refine curriculum and instruction and to target students who need the most help. Conversely, they might encourage narrow rote learning and the neglect of students perceived to have no chance at passing. Students might be motivated to study or discouraged from attempting a test they believe they cannot pass. All of this can mean a higher-quality graduating class with more valuable diplomas, but it may come at a cost of lower graduation rates—especially among students who do not speak English as their first language or who teachers may be more likely to discount. While exit exams will always need refinement to ensure that they adequately measure appropriate standards, the first concern with the tests at a policy level is equity, for which the first indication would be a significant drop in overall graduation rates.

3.3 Previous Findings on Exit Exams

Prior studies of HSEEs have examined their effects on dropout, completion, and graduation rates, with some additional work on their impacts on achievement and student educational trajectories.

More recent studies of HSEEs and dropout emphasize heterogeneous effects across student groups and the potential for exams to increase inequity. HSEEs tend to increase dropout rates, especially in disadvantaged populations (Bishop and Mane, 2001; Dee and Jacob, 2006; Jacob, 2001). The average effects of educational policies can be misleading as they often have distributional effects (Bitler et al., 2006). This is highlighted by Ou (2010), who uses a regression discontinuity design around the margin of barely failing and passing the New Jersey HSEE to find that while barely failing students are generally more likely to drop out than barely passing students, the negative effect is strongest for minority and low-income students. Dee and Jacob (2006) find that HSEEs reduce the probability of completion overall, but that these effects come from black students and school districts with high levels of poverty or minority enrollments in urban areas, while the HSEE actually lowers dropout rates in more affluent districts. HSEEs have the potential to enhance attainment by focusing instruction and student effort, but only in the presence of adequate resources. More critically, these potential positive effects are apparently weaker than the systematic negative effects of poverty and minority status in American education (Plunk et al., 2014).

HSEEs themselves can shape the educational trajectories of students. Evidence from a study on Turkish data indicates that high-stakes examinations may actually help reduce achievement gaps based on student background by promoting learning over the course of multiple re-takings (Frisancho et al., 2013). This supports the intuition that HSEEs may motivate students and help focus their efforts on mastery of the relevant material, but also highlights the importance of designing an exam that measures relevant content. Reardon et al. (2010) examine the impact of failing an HSEE in 10$^{\text{th}}$ grade—with two years remaining to pass before graduation—and find that barely failing the exam has no effect on students' academic trajectories, course taking, or graduation probability except for the very lowest achievers. Those authors conclude that negative effects of HSEEs on graduation rates come exclusively from the very lowest achievers. This supports the assertion that HSEEs may have overall positive effects, precluding from graduation only those students who may not be prepared to earn a diploma at all. Still, this does not address the equity issues that have been empirically demonstrated in the racial and socioeconomic distributions of these effects.

Graduation rates can be a more reliable measure of how many students are successfully completing their secondary education than dropout rates as they are more difficult to manipulate and there are fewer incentives to do so. For graduation rates, the effects of HSEEs are even more likely to be null (Carnoy and Loeb, 2002; Greene and Winters, 2005; Grodsky et al., 2009; Warren and Edwards, 2005), again with some exceptions (Amrein and Berliner, 2002; Marchant and Paulson, 2005). In strictly causal analyses, there appears to be a slight increase in dropout rates among black and low-income students when HSEEs are in place (Dee and Jacob, 2006; Murnane, 2013; Warren et al., 2006). Overall, there is some evidence that HSEEs can improve student achievement and the attainment of high school diplomas, but that this is highly dependent on school resources and subject to major racial and socioeconomic equity issues.

Graduation rates in general have been steadily improving throughout the first decade of the 21st century, especially among certain student groups. In a review of the topic, Murnane (2013) outlines the rise of graduation rates through the 1900s until their stagnation in the last three decades of that century, then their recent improvement. The recent rise in graduation rates appears to come largely from major increases in high school graduation among black and Hispanic students, an increase that has occurred simultaneously with the implementation of HSEE policies in a number of states. It remains unclear why graduation rates initially stagnated or restarted, and significant gaps based on race, gender, and socioeconomic status still exist. The role of HSEEs in these trends is also unknown.

The impact of HSEEs on graduation rates remains difficult to determine, especially in the long term. This is especially difficult because of the lack of controlled experimental conditions; counterfactual conditions and randomization simply do not exist. Regression discontinuity designs around the margin of barely passing have been very useful in identifying the effects of exams for students' trajectories and graduation probabilities (Dee and Jacob, 2006; Reardon et al., 2010), but overall trends and effects are difficult to determine. We examine long-term trends in graduation rates surrounding and following the year in which HSEEs were first used to withhold diplomas. By by exploiting the temporal variation in state-level implementation of HSEEs, we are able to mitigate year- and state-specific trends and show a broader and more generally applicable picture.

3.4 Data and Descriptive Statistics

In this section, we briefly describe our data collection process and present descriptive statistics. We combine data from multiple sources into a unique data set. The main source of information is the Center on Education Policy (CEP), which gathers state reports on high school education and high school examination procedures. From CEP's state reports we know whether a given state has an HSEE, when it was first administered (or reformed), and how it is structured in terms of grade alignment and content. We also double-checked our information with that of Dee and Jacob (2006), finding almost no difference.

To construct a panel, we complemented the data from CEP with information on high school graduation rates and achievement scores. We took graduation rates for the period between 1990 and 2012 from the Digest of Education Statistics (2012), produced by the National Center for Educational Statistics (NCES). We completed the NCES graduation rate data with information from America's Health Ranking,[2] which has prepared annual reports on health and health dynamics since 1990 that include high school graduation rates for each state from 1990 to 2013.

For achievement, our other outcome of interest, we rely on the NCES, which is also responsible for gathering data on the National Assessment of Educational Progress (NAEP). The NAEP is the most representative long-term assessment of the skills and

[2] http://www.americashealthrankings.org/

Table 3.1: Descriptive Statistics

	Mean [1]	Std. Dev. [2]	Min [3]	Max [4]
Panel A. All States (50 States, $N = 1,200$)				
Year	2001.50	6.93	1990	2013
HSEE	0.56	0.50	0.0	1.0
Graduation Rate	73.89	8.40	48.0	91.6
NAEP Score (8th Grade Math)	277.02	10.30	246.0	301.0
Panel B. States with HSEE (28 States, $N = 672$)				
Year	2001.50	6.93	1990	2013
First HSEE Administered	1999.86	5.53	1990	2012
Graduation Rate	70.24	8.31	48.0	89.7
NAEP Score (th Grade Math)	274.54	10.85	246.0	301.0

Notes: Data collected by the authors.

abilities of American students, and is reported at the state level in 4th and 8th grades. Assessments are not conducted every year, but still represent the best source for comparable state-level achievement data; we use 8th grade math scores as our measure of achievement. NAEP scores are a well-known measure of achievement in the research community.[3] Altogether, the resulting data forms a longitudinal state-level aggregated data set covering the period 1990-2013. The panel is balanced for the outcome "graduation rate" but not for the outcome "achievement," because NAEP scores are not collected every year.

Table 3.1 presents descriptive statistics, shown separately for all states (Panel A) and only for those that introduced an HSEE (Panel B). For a detailed table with HSEE introduction years for each state, see appendix Table B.3. Table 3.1 shows that 56 percent of states—28 in total—have introduced an HSEE.[4] Almost all states that have an HSEE introduced it between 1990 and 2012, excepting Alabama, New York, and South Carolina. These three states had an HSEE before 1990,[5] but reformed it in the 1990-2012 period. In our econometric analysis, we consider the reform year as the year of HSEE introduction for these three states.

Table 3.1 also presents descriptive statistics for our outcomes of interest. Graduation rates are a topic of much discussion in the HSEE literature because they are the area

[3] See Dee et al. (1999)

[4] We exclude the District of Columbia from our entire analysis.

[5] Alabama introduced an HSEE in 1984, New York in 1878 (Regents examination), and South Carolina in 1986.

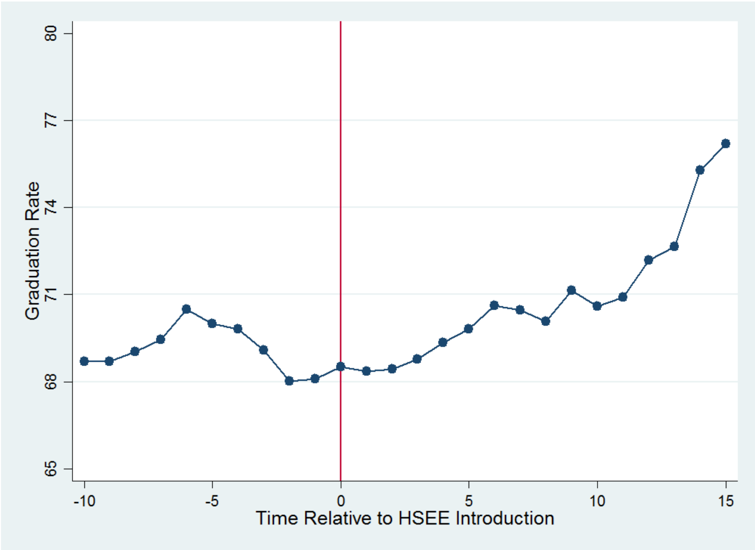
where the effects of HSEEs are most obvious and measurable. Most of this literature focuses on individual states in the short term following the application of an exam; we investigate graduation rates over the longer term and across all states with HSEEs. The average graduation rate for all states in the period 1990-2013 is about 74 percent, with a minimum of 48 percent in South Carolina (in 2003) and a maximum of 91.6 in Vermont (in 1992). The average graduation rate for the states that introduced an HSEE is almost four percentage points below the national average, which begs the question of whether their introducing an HSEE somehow helped the states without an HSEE to catch up; we answer this question in our robustness checks.

The effects of HSEEs for achievement and the quality of education is a more challenging question, especially for graduates' achievement following high school. Part of the intended effect of HSEEs on education quality is their ability to focus student, teacher, and school effort on ensuring students' knowledge of minimum requirements, so we assume that the implementation of an HSEE would trigger system-wide efforts to increase attainment. For this reason and to avoid any potential effects of dropout in 12th grade scores, we use 8th grade NAEP mathematics scores to represent the level of achievement in each state's educational system as a whole. We choose mathematics because state standards are clearest and most nationally consistent on that subject. The national average NAEP score for 8th grade math is 277, whereas the average for states that introduced an HSEE is 274.5. Because NAEP scores are not collected annually, we only have 457 observations for this outcome (260 for the sub-sample of states with an HSEE). This limits the potential significance of our results but does not compromise the integrity of our analysis as an indication of trends in achievement.

3.5 Empirical Strategy

State-level time trends in high school graduation rates and achievement are a natural point of departure for considering the impact of introducing an HSEE on these two outcomes. Figure 3.1 and Figure 3.2 present trends in state-level graduation rates and

Figure 3.1: Average Graduation Rate over Time Relative to HSEE Introduction (28 States)

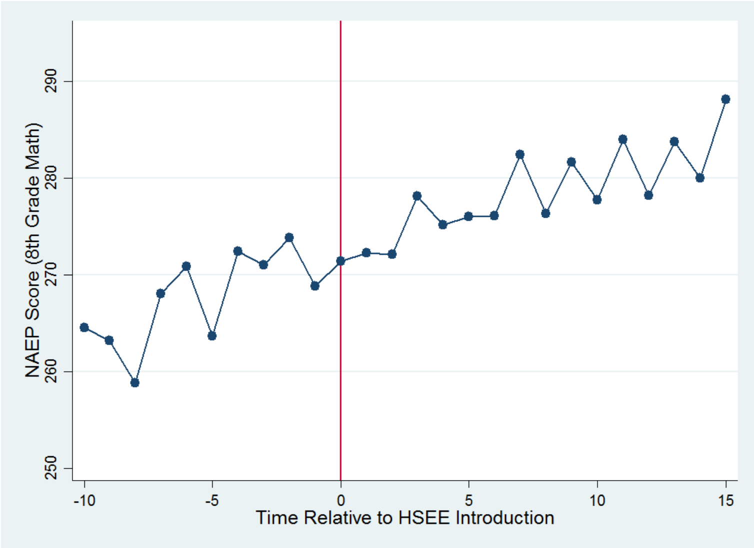


eighth grade NAEP math scores from 1990 to 2013. The vertical line visually identifies the relative point in time when an HSEE was implemented. The trends shown in Figures 3.1 and 3.2 suggest that introducing an HSEE may have had some positive effects on graduation rates and eighth-grade math scores, at least for those states that did introduce one. However, potential state- or nationwide changes in educational, social, and economic factors over this period make it difficult to credibly identify causal inferences from these trends.

To circumvent these concerns, we use an interrupted time series approach. The basic intuition of interrupted time series designs is to use pre-treatment trends in the outcome as a counterfactual—what the post-treatment trend in the outcome would have been in the absence of the treatment. Given that this assumption might be overly strong in some cases—especially if other factors have changed together with the treatment—we perform a robustness check using states that do not have an HSEE as control group. The interrupted time series approach has a long tradition in education research,[6] and has been used recently to evaluate several educational reforms such as the No Child Left Behind act (Dee and Jacob, 2011; Dee et al., 2012) and Accelerated Schools (Bloom et al., 2001).

[6]For an overview, see Shadish et al. (2002).

Figure 3.2: Average Achievement over Time Relative to HSEE Introduction (28 States)



In general, we can describe the trend of an outcome Y in state i at time t as:

$$Y_{it} = f(t - t_i^*) + HSEE_{it} \cdot g(t - t_i^*) + \gamma_i + \varepsilon_{it} \tag{3.1}$$

where $HSEE_{it}$ indicates the treatment status of unit i at time t, which in our case is an indicator of whether the state has an HSEE at time t. Time is measured in both continuous (t) and relative (t_i^*) metrics. Specifically, t_i^* denotes the time at which treatment begins in unit i; therefore t_i^* is the period such that $HSEE_{it} = 1$ if $t \geq t_i^*$ and $HSEE_{it} = 0$ if $t < t_i^*$. Subscript i in in the relative metric of time is necessary because states introduced their HSEEs at different points in time. This strengthens our causal claims by filtering out other changes or policies that may have happened concurrently. The function f describes the trend prior to t_i^*, and starting from period t_i^* the trend in Y_{it} is described by the function $(f + g)$. Under the assumption that the trend described by f would have continued after t_i^* in the absence of treatment, the effect of $HSEE_{it}$ by time $t \geq t_i^*$ is given by $g(t - t_i^*)$. Finally, γ_i represents state fixed effects and ε_{it} is an error term.

In interrupted time series designs, multiple effects exist. First, we can estimate a sharp discontinuity at the time of intervention—a change in level. Second, we are able to estimate the change in the slope of the time series at the point of intervention—a kink point. Third, we can estimate a continuous (or discontinuous) effect that does not (or does) decay over time. Fourth, it is possible to study potential intervention effects that are immediate, delayed, or even anticipatory. Depending on the functional form we impose on the functions f and g in Equation 3.1, we can focus on any effects of interest. Most simply, we might approximate f and g as linear functions of time as follows:

$$Y_{it} = \beta_0 + \beta_1 \cdot (t - t_i^*) + \beta_2 \cdot HSEE_{it} + \beta_3 \cdot (t - t_i^*) + \gamma_i + \varepsilon_{it} \tag{3.2}$$

The model specified in Equation 3.2 says that the trend in Y_{it} before time t_i^* is linear with slope β_1. At time t_i^*, the value of Y_{it} changes by β_2, then the trend in Y_{it} after time t_i^* is linear with slope $(\beta_1 + \beta_3)$. The effect of introducing an HSEE by time $t \geq t_i^*$ is given by $\beta_2 + \beta_3 \cdot (t - t_i^*)$. We can test the null hypothesis that the effect at time t is zero with the following F-test: $\beta_2 + \beta_3 \cdot (t - t_i^*) = 0$. In our linear specification, β_2 can be seen as a regression discontinuity estimate of the immediate effect of HSEEit on Y_{it}, with the difference that in interrupted time series we observe one unit at different points in time whereas in a regression-discontinuity design we would observe multiple units at one point in time. Similarly, β_3 can be seen as a difference-in-differences estimate of the effect of HSEEs on Y_{it}, or the difference in the rate of change for the average Y_{it} between treated and untreated states.

As we explained, Equation 3.2 allows us to estimate two effects, namely the change in level and the change in slope at the time of intervention. However, the treatment effect might decay or reinforce itself over time, and in cases like these our linear specification would lose the pattern. To allow the effects of the introduction of an HSEE to be a non-parametric function of time, we specify the following model:

$$Y_{it} = \beta_0 + \beta_1 \cdot (t - t_i^*) + \sum_{j=0}^{J} \delta_j \cdot D_t^j + \gamma_i + \varepsilon_{it} \tag{3.3}$$

where is a dummy variable equal to one if $(t - t_i^*) = j$, and J is the number of years observed after t_i^* (in our case $J = 15$). In Equation 3.3, the effect of $HSEE_{it}$ on Y_{it} j years after the introduction of an HSEE is represented by δ_j, and we can test the null hypothesis of no treatment effect at year j by testing $H_0 : \delta_j = 0$.

As we discussed before, another interesting effect that can be seen in in interrupted time series models is the anticipation effect, or whether the intervention has an effect on the outcome before it is actually introduced. To find this, we allow the effect of introducing an HSEE to be a fully non-parametric function of time:

$$Y_{it} = \beta_0 + \sum_{j=-10}^{J} \delta_j \cdot D_t^j + \gamma_i + \varepsilon_{it} \tag{3.4}$$

where D_t^j is, again, a dummy variable equal one if $(t - t_i^*) = j$, and J is the number of years observed after t_i^* ($J = 15$). In this specification, j can also be negative to estimate the trend in Y_{it} for each pre-treatment period up to ten years before the introduction of an HSEE. The non-parametric approach we use in Equation 3.4 is well known and used among labor economists, for example in estimating wage losses after job separation.[7]

In time series settings, estimating pre-intervention effects non-parametrically also constitutes an important test for causality. As suggested in the early literature by Granger (1969, 1988), the availability of many years of data enables testing of whether changes in a given policy lead or lag the outcome. If the within-state changes in Y_{it} lag—or coincide with—the within-state intervention, then they are consistent with a causal story in which causes are followed by their effects. In contrast, if within-state changes in the outcome lead the within-state changes in the policy, we might normally suspect either policy endogeneity or the presence of unobserved time-varying state characteristics that drive the effects. In our setting, however, it is hard to believe that introducing an HSEE had no pre-intervention effects, especially because such a policy follows a lengthy political discussion involving schools, districts, and the state government. Therefore, we might expect an effect of the HSEE on graduation rates or even achievement a few years before its actual introduction. For example, if schools know that an exit exam is going to be introduced soon, they might modify their current high school curricula in anticipation to ensure that students are prepared.

[7]See, for example, Jacobson et al. (1993); von Wachter et al. (2008); Balestra and Backes-Gellner (2012).

In sum, to estimate the effect of introducing an HSEE on graduation rates and achievement, we use an interrupted time series approach and three different specifications (Equations 3.2, 3.3 and 3.4). We rely on different specifications because we are interested in the many effects an HSEE might have. While the linear specification (Equation 3.2) focuses on the changes in level and slope at the time of intervention, the semi-parametric (Equation 3.3) and fully non-parametric (Equation 3.4) specifications allow the estimation of treatment effects specific to each year before, during, and after the intervention.

3.6 Results

This section presents our results in three parts. The first subsection shows the results for the graduation rate outcome, starting from the linear specification (Equation 3.2) and then presenting our semi-parametric (Equation 3.3) and non-parametric specifications (Equation 3.4). The second subsection examines the effect of introducing an HSEE on achievement. In the third subsection, we perform robustness checks to test the internal and external validity our empirical strategy.

3.6.1 Effect of HSEE on Graduation Rates

Table 3.2 presents regression outputs for the linear specification in Equation 3.2. We estimate three effects: the trend in graduation rates before the introduction of an HSEE ($t - t_i^*$), the change in level of graduation rates in the year of introduction ($HSEE$), and the change in slope after the introduction of an HSEE (the interaction term). Note that we are estimating a within-state effect for the states that introduced an HSEE, and thus the counterfactual is the state's pre-intervention trend in graduation rates.[8]

From Table 3.2 we infer that there is no particular trend in graduation rates before the introduction of an HSEE, because the coefficient of $(t - t_i^*)$ is not significant. Similarly, there is no significant discontinuity at the time of intervention. However, and most importantly for our research question, we estimate a positive change in slope for the time

[8]We consider alternative counterfactuals in subsection 5.3, in which we perform our robustness checks.

Table 3.2: Effect of HSEE on Graduation Rate, Linear Specification

Variables	Graduation Rate Coefficient [1]	Standard Error [2]
$(t - t_i^*)$	-0.141	(0.118)
HSEE	-0.871	(1.129)
$(t - t_i^*)\cdot$ HSEE	0.562***	(0.144)
Intercept	68.447***	(0.885)
State Fixed Effects	YES	
Adjusted ^{2}	0.777	
N	672	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

series at the time of intervention. This kink point is significant at the highest confidence level (p-value = 0.00), and means that introducing an HSEE has a positive impact on the trend in graduation rates.

In sum, according to our linear specification, the within-state graduation rate trend before introducing an HSEE is rather stable, then becomes positive after implementation. This positive pattern is consistent with previous research on HSEEs for industrialized countries (Bishop et al., 2001). However, by specifying a linear function of time we are assuming that the effect of introducing an HSEE is constant and are not allowing for heterogeneous effects over time. We might suspect that this assumption is invalid if the immediate effect decays or reinforces over time. Therefore, we also estimated semi- and non-parametric models.

Table 3.3 shows the results of the semi-parametric (columns 1-2) and non-parametric (columns 3-4) specifications. The semi-parametric model assumes a linear trend in graduation rates before HSEE introduction and a separate effect in each year thereafter. Although we know from Table 3.2 that the overall trend in graduation rates becomes positive after introducing an HSEE, the semi-parametric specification reveals that the trend is initially negative. For the first four years with a new HSEE, within-state graduation rates decrease by almost three percentage points. This loss is statistically significant and negative in the first four years, at which point it becomes insignificant.

A similar picture emerges from the non-parametric specification, which estimates the time series as a non-parametric function of time with the treatment effect estimated separately for each year before and after HSEE introduction. In the third column of Table 3.3, we observe that the significant short-term loss lasts up to three years after introducing an HSEE and has a magnitude of slightly more than three percentage points per year. Note that, in the non-parametric specification, the estimated effects are relative to the year of introduction.[9] In the long term, the effect of introducing an HSEE eventually becomes positive and marginally significant, which drives the results of the linear specification.

In the non-parametric specification, we can study not only effects following the introduction of an HSEE but also any potential anticipatory effects. As discussed in Section 4, we might expect states to adjust their curricula before the actual introduction of an HSEE in order to prepare their students, teachers, and schools for the new HSEE requirement. Furthermore, major educational policy changes like HSEEs are usually accompanied by several years of political and popular discussion and we can easily assume that students, teachers, and schools were aware of the impending introduction of an exam. The results in column 3 of Table 3.3 confirm our expectations, with statistically significant pre-intervention effects one and two years before HSEE introduction. The presence of pre-intervention effects, however, might weaken our causal claims because it might appear that the effects are leading—rather than lagging—the cause. Still, the availability of many years of data and the fact that stakeholders are aware of HSEEs before their introduction reinforces our interpretation of the results.

In sum, we find that introducing an HSEE has an overall positive effect on graduation rates and a positive effect on the slope of the time series for graduation rates. However, this effect on graduation rates is heterogeneous over time. In the year of introduction and for at least the following three years, HSEEs have a negative impact on graduation rates. This negative impact is short-lived and becomes positive towards the end of our timespan. We also estimate a pre-intervention negative effect one and two years before HSEE introduction, suggesting that students, teachers, and schools start preparing for

[9] Changing the base category (e.g., setting the first year as base category) has no impact on the results.

Table 3.3: Effect of HSEE on Graduation Rate, Non-Parametric Specifications

Variables	Graduation Rate			
	Coefficient	Standard Error	Coefficient	Standard Error
	[1]	[2]	[3]	[4]
$(t - t_i^*)$	0.198*	(0.072)		
$(t_i^* - 10)$			-1.710	(1.145)
$(t_i^* - 9)$			-2.020	(1.162)
$(t_i^* - 8)$			-2.045	(1.181)
$(t_i^* - 7)$			-1.793	(1.246)
$(t_i^* - 6)$			-1.433	(1.251)
$(t_i^* - 5)$			-1.638	(1.183)
$(t_i^* - 4)$			-1.834	(1.289)
$(t_i^* - 3)$			-2.551	(1.257)
$(t_i^* - 2)$			-3.617*	(1.337)
$(t_i^* - 1)$			-3.644**	(1.087)
(t_i^*)	-2.448**	(0.663)	*Base Category*	
$(t_i^* + 1)$	-2.785**	(0.774)	-3.379**	(1.040)
$(t_i^* + 2)$	-2.723**	(0.937)	-3.193**	(1.100)
$(t_i^* + 3)$	-2.732*	(1.084)	-3.076*	(1.217)
$(t_i^* + 4)$	-2.342	(1.276)	-2.487	(1.451)
$(t_i^* + 5)$	-1.976	(1.235)	-1.985	(1.364)
$(t_i^* + 6)$	-1.374	(1.410)	-1.185	(1.545)
$(t_i^* + 7)$	-1.712	(1.520)	-1.325	(1.665)
$(t_i^* + 8)$	-2.371	(1.655)	-1.828	(1.768)
$(t_i^* + 9)$	-1.988	(1.814)	-1.285	(2.074)
$(t_i^* + 10)$	-2.271	(1.623)	-1.405	(1.814)
$(t_i^* + 11)$	-2.165	(1.770)	-1.101	(1.902)
$(t_i^* + 12)$	-1.068	(1.978)	0.195	(2.058)
$(t_i^* + 13)$	-0.758	(1.847)	0.547	(1.823)
$(t_i^* + 14)$	0.330	(1.248)	1.747	(1.231)
$(t_i^* + 15)$	1.064	(1.302)	2.606	(1.304)
Intercept	70.951***	(0.515)	71.743***	(0.946)
State FE	YES		YES	
Adjusted R^2	0.760		0.749	
N	672		672	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

exams even before their actual introduction.

3.6.2 Effect of HSEE on Achievement

Our second outcome of interest is achievement. The only measure of achievement we have for all states and for multiple years is the NAEP score. As many other studies do Dee and Jacob (2011), we rely on 8^{th} grade math scores as a measure for achievement. We choose math because mathematical skills are relatively easier to measure through standardized tests, state standards for mathematics are most consistent, and the potential for bias is minimized. Similarly, we choose 8^{th} grade because we want to analyze a grade as close as possible to high school without introducing the possibility of selection bias from dropout.

We believe that a introducing an HSEE impacts the behavior of students and teachers and the development and application of curriculum not only at the high school level but throughout a state's education system. The general trend in education policy in the United States over the past decades supports this intuition: HSEEs themselves are part of a broader trend towards streamlining, quantifying, and building accountability in education. As is especially obvious following the introduction of the Common Core State Standards, states are attempting to set and assess clear and rigorous standards at all grade levels.

Table 3.4 presents regression outputs for the linear specification. The number of observations is halved because we have an unbalanced panel for the achievement outcome. The only significant coefficient in Table 3.4—aside from the intercept—is the linear trend in time, which is positive and highly significant. Therefore, it appears that introducing an HSEE has no impact on the (positive) trend in achievement: we estimate neither a discontinuity nor a kink.

Table 3.5 shows estimated effects of HSEEs on achievement for the alternative semi- and non-parametric specifications. As in Table 3.4, we estimate no particular effect of introducing an HSEE on achievement. We nevertheless observe a positive trend in our non-parametric specification—a trend that becomes statistically significant about ten years after HSEE introduction. Note that the multitude of non-significant coefficients in

Table 3.4: Effect of HSEE on Achievement, Linear Specification

Variables	NAEP Scores (8^{th} Grade Math) Coefficient [1]	Standard Error [2]
$(t - t_i^*)$	0.912***	(0.159)
HSEE	0.627	(1.009)
$(t - t_i^*)\cdot$ HSEE	0.083	(0.180)
Intercept	270.562***	(0.952)
State Fixed Effects	YES	
Adjusted ^{2}	0.927	
N	260	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

the non-parametric specification is likely due to the relatively small sample size and the many degrees of freedom we lose by estimating all the parameters.

We conclude that introducing an HSEE has no statistically significant impact on the time series of achievement. This result holds for the states that introduced an HSEE, which have lower achievement on average compared to those that never had an HSEE. In the next subsection we perform the analysis including states that have no HSEE as part of the control group. This not only constitutes a robustness check for our results, but also shows whether introducing an HSEE helps to close the achievement gap between states with HSEEs and states without by raising the level of the states that introduced HSEEs from much lower than their non-HSEE counterparts to only slightly lower.

3.6.3 Robustness Checks

In this subsection, we estimate our models for the full set of all states. Doing so allows us to use states without HSEEs as counterfactuals for those that have one. Including all of the states in the regressions has two purposes. First, we test the robustness of our results. If we find completely different effects, it would mean that states with HSEEs are completely different from those without HSEEs, casting some doubt on the external validity of our results. However, if the results are in fact consistent with those of the previous subsections, it would mean that states are rather similar among themselves—at least af-

Table 3.5: Effect of HSEE on Achievement, Non-Parametric Specifications

Variables	NAEP Scores (8^{th} Grade Math)			
	Coefficient	Standard Error	Coefficient	Standard Error
	[1]	[2]	[3]	[4]
$(t - t_i^*)$	0.994***	(0.066)		
$(t_i^* - 10)$			-9.300*	(4.111)
$(t_i^* - 9)$			-7.257	(3.720)
$(t_i^* - 8)$			-8.502*	(3.796)
$(t_i^* - 7)$			-5.884	(4.620)
$(t_i^* - 6)$			-5.366	(4.034)
$(t_i^* - 5)$			-5.266	(3.775)
$(t_i^* - 4)$			-0.776	(3.202)
$(t_i^* - 3)$			-0.650	(5.150)
$(t_i^* - 2)$			-0.874	(3.849)
$(t_i^* - 1)$			-2.563	(4.626)
(t_i^*)	-0.822	(1.730)	*Base Category*	
$(t_i^* + 1)$	0.510	(1.001)	-0.198	(5.914)
$(t_i^* + 2)$	-1.187	(1.025)	0.321	(4.041)
$(t_i^* + 3)$	2.636*	(0.979)	5.981	(3.624)
$(t_i^* + 4)$	-0.754	(1.141)	3.465	(3.512)
$(t_i^* + 5)$	1.319	(0.916)	5.169	(4.330)
$(t_i^* + 6)$	0.520	(1.160)	5.577	(4.195)
$(t_i^* + 7)$	1.513	(1.004)	8.463*	(3.490)
$(t_i^* + 8)$	-0.368	(1.203)	7.050	(3.961)
$(t_i^* + 9)$	0.978	(1.639)	8.611	(4.283)
$(t_i^* + 10)$	-0.054	(1.162)	7.836	(4.200)
$(t_i^* + 11)$	0.680	(1.740)	10.809**	(3.700)
$(t_i^* + 12)$	-1.408	(1.145)	8.402	(4.319)
$(t_i^* + 13)$	-0.221	(1.576)	10.004*	(3.945)
$(t_i^* + 14)$	-1.778	(1.094)	7.412	(4.268)
$(t_i^* + 15)$	0.234	(1.182)	11.218**	(4.038)
Intercept	271.185***	(0.412)	272.067***	(3.036)
State FE	YES		YES	
Adjusted R^2	0.927		0.636	
N	260		260	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

Table 3.6: Effect of HSEE on Graduation Rate and Achievement, All States

Variables	Graduation Rate		NAEP Scores (8^{th} Grade Math)	
	Coefficient	Standard Error	Coefficient	Standard Error
	[1]	[2]	[3]	[4]
$(t - t_i^*)$	0.033	(0.045)	0.752***	(0.061)
HSEE	-1.946*	(0.936)	1.670	(0.903)
$(t - t_i^*)\cdot$ HSEE	0.395***	(0.094)	0.234*	(0.100)
Intercept	73.335***	(0.296)	269.791***	(0.348)
State Fixed Effects	Yes		Yes	
Adjusted R^2	0.799		0.921	
N	1,200		457	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

ter controlling for time-invariant characteristics. This might increase the generalizability of our results to all states. Second, using states without HSEEs as counterfactuals might have relevant policy implications. Given that states with HSEEs have lower graduation rates and lower achievement, we can investigate whether introducing an HSEE helps close the graduation rate and achievement gaps between states.

Table 3.6 shows regression outputs for the effect of HSEEs on graduation rate (columns 1-2) and achievement (columns 3-4). We restrict our analysis to the linear specification for easy comprehension and because the semi- and non-parametric models—which are presented in appendix Tables B.1 and B.2—reveal no additional information. First and foremost, we observe that the effects are very similar to those of the previous tables, especially in terms of direction. However, we find some differences in terms of significance, which might be partly due to the increase in sample size.

One relevant finding of Table 3.6 is that introducing an HSEE appears to help close both graduation rate and achievement gaps between states with HSEEs and those without. For graduation rates, we estimate a negative and significant discontinuity at the time of HSEE introduction of about two percentage points. We also estimate a highly significant change in the slope of the graduation rate time series, which mirrors the kind we found earlier. Regarding achievement, we estimate the same highly significant and positive long-term trend as before. Moreover, we also find a positive and significant change in slope for those states that introduced an HSEE after its introduction. HSEEs alone may not cause these increasingly positive trends in graduation rates and achievement, but some part of the streamlining and focusing of educational standards that includes HSEEs and policy does cause them.

Overall, we conclude that our results are not sensitive to the control group chosen. Additionally, we find that introducing an HSEE helps reduce the graduation rate and achievement gap between states with HSEEs (usually lower performing) and states without HSEEs (usually higher performing).

3.7 Conclusions and Discussion

There are some potential mechanisms behind the improvement in graduation rate trends and—in some specifications—the similar improvement in achievement growth following the implementation of an HSEE. The first possibility is curricular improvement and enhanced behavior of schools, teachers, and students. This is the intended effect of HSEEs, and the increased accountability from the exam could cause such improvements. Second, the focusing of curriculum, instruction, and student effort could be responding to the perverse incentive to narrow effort down to only the material covered on the exam. In the case of an HSEE that perfectly represents and measures the skills and knowledge entailed by a high school diploma, this would be identical to the first possibility. If there are design or measurement issues with the exam, however, this would place limitations on the value and effectiveness of students' education. In both of these cases, the improvement in graduation rates and math test achievement we find would be explained by the HSEE successfully modifying educational behavior towards increased achievement—at least as measured by the exam itself.

Alternatively, changes in the HSEE such that the exam adapts to a state's students rather than the other way around could yield the increase in graduation rates observed in this study; instead of the HSEE modifying behavior as intended, revisions of the exam due to political or social pressure may improve students' likelihood of passing. One possibility for such an adjustment—one that has been called for and implemented frequently—is an option for non-native English speakers to take the exam in their native language. HSEE policies have been accompanied in many states by an outcry against potential discrimination against immigrant students, and most of those states have added concessions for non-English speakers. These concessions increase the passing rate for the groups of students they target, which would in turn raise the overall passing rate. A second possible adjustment would be to simply lower the HSEE's difficulty level or adjust its format in response to claims that it was too rigorous or discriminatory. Many states are moving towards end-of-course rather than comprehensive exams, and the content of exams changes frequently. These changes may also contribute to higher passage rates on average.

Over the long term, HSEEs do not appear to decrease graduation rates, instead they strengthen improvement trends in graduation rates. More importantly, this increase in graduation rates is not accompanied by any decrease in achievement—more students are graduating and meeting the increased standard. Graduation rates dip in the short term immediately before, during, and after the year in which exam scores are first used to withhold diplomas, but recover soon after and eventually improve over pre-HSEE trends. The mechanism is still unclear; while improvements may come from the adjustments in school, teacher, and student behavior as intended by policy makers, improvements in graduation rates may also come from narrowing of teaching and learning to exam material or changes in the exam itself rather than student achievement. Further research on how HSEEs interact with graduation rates and especially achievement is necessary, but the first step has been to understand how they affect graduation rates and achievement overall.

It is true that students at the margin of failing the exam whose graduation year is on or near the first year of an exit exam policy will have decreased probabilities of graduation, but these effects are not persistent and are counterbalanced by later improvements in student attainment. If exams succeed in modifying school, teacher, and student behavior towards a more focused mastery of the material required to graduate high school, the exams are useful insofar as they accurately reflect and measure such material. These findings may help rectify the lack of convergence in the literature towards a standard result for HSEEs and graduation rates; the net effect of HSEEs for graduation rates is positive over the long term, but shorter-term effects can be negative. By estimating both effects, we rectify some of the disagreement in the literature emerging from differences in data sources and estimation strategies. The simultaneous effects of raising attainment by enhancing behavior and improving the signaling value of a high school diploma are important steps forward for secondary education.

Chapter 4

Predicting School Achievement Reactions to Increased Funding

4.1 Introduction

School funding[1] is one of the most challenging policy issues: everyone agrees that it is important but neither public opinion nor educational researchers agree on the best approach. Education is a major part of government spending in developed countries (OECD, 2013), but the link between educational expenditure and student achievement is tenuous at best (Hanushek et al., 2011). Methodological biases and limitations are commonly cited as contributing to the disagreement and lack of evidence on the relationship that theoretically exists between funding and outcomes in education. In this paper, I take a new approach from the rising field of big data analytics and employ a model-free prediction technique rather than more common model-based exploration methods. I find that predicted outcomes for the highest- and lowest-achieving schools do not change with additional funding, while there is great variation in predicted outcome change among schools in the middle of the distribution. Schools in the third quartile of the distribution are most likely to have improved predicted outcomes with more funding.

Australian education is already on the leading edge of data usage in education.

[1]This chapter is written entirely by Katherine Caves. It has been submitted to the 2015 World Educational Research Association Conference.

According to Pugh and Foster (2014), big data in education is on the rise and the Australian Curriculum, Assessment, and Reporting Authority (ACARA) has the potential to lead the way. In their words, "education research is far more exciting today that in our parents' day, due largely to the Big Data revolution in which Australia is already playing an important role, as a small country with national testing and an organised means of disseminating some of its education data to researchers." Not only is the use of large datasets and their accompanying analytical techniques increasingly relevant for educational policy research, Australia is taking steps towards ensuring that their data infrastructure can support new forms of analysis. Making educational data available to researchers and encouraging the use of educational data mining—a class of computer-driven statistical learning techniques—can help educational policy makers gain insight into their existing system and develop new systems.

Educational data mining techniques for "classification, clustering, association rule mining, sequential mining, text mining" (Romero and Ventura, 2010) and other purposes have been used in educational research, but their application remains limited. They have mostly been applied to online learning in the form of Massive Open Online Courses, and in online or computer-based educational support software. Up to this point, the kind of data required for these techniques has been difficulty to acquire. Educational data mining techniques are most often used at the student-to-instructor level instead of organizational or administrative levels. In their review of the educational data mining field, Romero and Ventura (2010) lay out potential avenues for future research at the student, teacher, curriculum development, organizational, and administrative levels. I take up their recommendations for administrative and organizational uses of educational data mining techniques. At the organizational level, educational data mining research can increase efficiency, help schools attain their state goals, and identify the most cost-effective means of raising scores and grades. At the administrative level, it can be useful for organizing and effectively using resources.

In this paper, I use administrative data from ACARA on every school in the state of Victoria, Australia. I use a common classification technique called k nearest neighbors (KNN) to predict the test score outcome of each school based on its background, demographic, previous achievement, and financial data. Subsequently, I assess the effects of additional funding on test score predictions along the predicted achievement distribution, identifying the patterns where additional funding is related to change in predicted achievement. I find that additional funding alone is not associated with large changes in school performance for schools at the lowest and highest ends of the predicted achievement distribution. Additional funding is associated with large changes in the predicted achievement in schools in the midrange of the predicted achievement distribution, but these changes are highly variable and not always positive.

4.2 Theory and Literature

The idea of intervening in educational performance through funding comes from the economic concept of educational production as a function of inputs and resources. Essentially, educational production—in this case achievement—should be a predictable outcome of the specific resources available to teachers, schools, and students and the quantities of those inputs. The Handbook of the Economics of Education summarizes the predicted relationship between funding and achievement as follows: "higher expenditure, all else equal, is typically assumed to deliver a higher quality of education" (Hanushek et al., 2011).

Many efforts to specify such a production function have been made, starting with the work of Hanushek (1986). Generally, findings in educational production are rather surprising: there is very little evidence that the positive impact of additional financial resources has any more than a very modest effect on achievement (Hanushek et al., 2011). Even in famous examples like Tennessee's Project STAR class size-reduction experiment, the initial finding of positive effects for some students of smaller classes have been heavily contested (see for example (Balestra and Backes-Gellner, 2014). Eventually, Hanushek concluded, "pure resource policies that do not change incentives are unlikely to be effective" (Hanushek, 2006, p. 865).

The conclusion that resources alone are unrelated to achievement is not universal. Krueger (2003) performs a meta-analysis of the educational production function literature and states, “when studies are given equal weight, resources are systematically related to student achievement” (p. F34). This opinion is not unique; other studies and reviews support this conclusion (see for example Hedges and Greenwald, 1996; Vergestegen and King, 1998; Card and Krueger, 1996; Nascimiento, 2008). Although there is little evidence for education funding as a driver of educational quality and achievement, the lack of evidence may be driven by limitations in methodology or data rather than the real absence of such a connection.

Public opinion and educational policy making tend to favor the second school of thought and often assume that additional resources are the key to educational improvement, even if the empirically demonstrated connection between funding and achievement is tenuous in that specific context. In Australia, the strongest predictor of school achievement is the socio-educational background of a school's community, and there is little or conflicting evidence that school funding affects achievement (Miller and Voon, 2011). Despite this, there is—as in many developed countries—a raging debate over the adequacy and appropriate use of education funding.

In general, Australian education policy tends toward using educational funding as a tool for improving student outcomes and equity. This is best summarized in the Gonski Report (Gonski et al., 2011), which argues that decreasing student performance and the growing achievement gap in Australia should be resolved by significantly increasing funding to schools. Increased funding is seen as the best course to improved outcomes. Schools may be desperate for adequate resources, but the aspirations of the report for the effects of increased funding go far beyond meeting basic needs. That report is not unique: the funding formula used by ACARA grants additional funding to schools with more disadvantaged students. Again, funding is conceived as a means for improving achievement beyond simply providing adequate education. In practice, it is difficult to resist the idea that increased educational funding will generate improved educational outcomes.

In this paper, I assess the validity of that claim using a novel methodological approach. The research debate has focused often on potentially biased or inadequate methodologies, so this is an attempt to add another perspective to the conversation. In addition, this methodology is something that educational administrators and decision-makers can use in their own contexts to guide policy choices. I use KNN to predict the achievement of schools in Victoria, then adjust their funding levels to identify patterns in reactions to funding changes along the achievement distribution.

4.3 Data

The data for this study is register data collected by the DEECD in Victoria on every government and non-government primary, combined, and secondary school in that state over three years from 2009-2012. After dropping special schools (schools for the deaf, etc.) and the very few schools without complete data, the sample includes 1843 schools for the school year ending in 2010, 1879 for the 2011 year, and 1893 in 2012. Schools are generally 34 percent non-government (Catholic and independent) and 66 percent government, and are approximately 73 percent primary, 11 percent combined, and 16 percent secondary. The descriptive statistics presented in Table 4.1 are for all three years of analysis aggregated together, and descriptive statistics for each year individually can be found in the appendix (Tables C.3, C.4, and C.5).

Data is divided into three categories. School profiles reflect the characteristics of each school. The Index of Community Socio-Economic Advantage (*ICSEA*) score is a measure calculated by ACARA to enable meaningful comparisons across schools. It includes census data and socio-economic data like average income, average education level, and typical employment types on the neighborhoods in which each school's students live. These scores are aggregated and reported on the school level and are standardized such that the national mean is 1000 points and the standard deviation is 100. Victoria is a relatively wealthier and more homogenous state, so its scores are slightly above the national average and the standard deviation is smaller.

Table 4.1: Descriptive Statistics for All Years Combined

N = 5,615	Mean	St. Dev.	Min	Max
Sector	1.842	0.558	1	3
Type	2.047	0.518	1	3
ICSEA Score	1,031.070	71.085	612	1,247
Full-Time Enrollments	422.452	385.525	7.000	3,289.000
Indigenous Enrollments	1.572	3.164	0	100
LBOTE	21.514	24.417	0	199
Teacher-Student Ratio	0.072	0.020	0.042	0.273
Staff-Student Ratio	0.024	0.014	0.002	0.210
Attendance Rate	92.984	2.862	50	100
Funding Per Student	11,018.920	4,122.437	2,012.738	68,152.060
Previous Score	482.980	50.591	312.500	725.000
Current Score	482.085	50.323	312.500	725.000

School profile data also includes information on the students in each school. *Full-time Enrollments* refers to school size by calculating the full time-equivalent number of students at each school—this prevents over- or under-counting as students change schools, enter, or drop out. *Indigenous Enrollments* is the percentage of each school's student body that is from native Australian ancestry. These numbers are quite low as Victoria is not a state with a large indigenous population. Students from a language background other than English (*LBOTE*) are also represented as a percentage of the student body, and represent immigrants and children of immigrants from non-English-speaking countries.

The last element of school profile data is information on the policies, hiring, and decisions made by each school's school council and principal. Schools in Victoria are quite autonomous, so these individual school policies have the potential to vary widely. *Teacher-Student Ratio* is the number of teachers per student in each school—it is a measure of class size. Like many developed countries, Australia faces a constant debate about whether educational policies aimed at reducing class sizes—some of the most expensive policy changes—are worthwhile. Similarly, *Staff/Student Ratio* captures the number of non-teaching staff members per student in each school. Finally, *Attendance Rate* captures the frequency with which students actually attend schools.

The second category of data is financial data. In this case, I have chosen not to disaggregate funding by source (federal government, state government, fee/parent funding, and other private funding) in order to focus on the broader context of finance in education rather than education finance in Australia. For this reason, funding is represented as *Funding Per Student* and is the number of Australian dollars at each school's disposal for each child.

Finally, the last category is achievement data. Each school is required to take the National Assessment of Progress—Literacy and Numeracy (NAPLAN), which is a standardized test of achievement administered in grades 3, 5, 7, and 9. Scores vary by testing year, but the analysis includes which grades each school serves so this is not an issue. *Previous Year Test Score* is historical data on the average score of each school in the year before the year at hand. *Current Year Test Score* is the outcome variable here, and reflects the test score achieved by each school in the year when its characteristics were measured.

This data set is the full population of schools rather than a sample, so there is no need to ensure its representativeness. For this study, I use all three waves of schools and their test score from the previous year to predict test scores in the year at hand. School profile, financial, and score data from 2010 is used with test scores from 2009 to predict scores in 2010, and so on for all three years. In this way I am able to cross-validate my analysis and compare the predicted test scores against real scores to determine the accuracy of the predictions.

Simultaneously, the use of one year's data to predict the next year's test scores reflects the realities of information available to educational decision-makers when they are attempting to predict which schools might need extra assistance in the next year. In this way, the analysis is as relevant as possible for education policy making.

4.4 Empirical Strategy

KNN was introduced in 1951 by Fix and Hodges (see Silverman and Jones, 1989) and formalized by Cover and Hart (1967). It is a "lazy" or instance-based learning method because it does not require a model, working instead directly on the training data. KNN can be used to predict the category of an observation or to predict its continuous value, depending on the research question. In the case of categorical prediction, the majority vote of similar observations is used to predict the test subject's category. In the case of continuous prediction, the average of the near neighbors' scores is used. The basic procedure is to divide the data into training and test sets, find the best k to minimize both mean squared error and overfitting, and use the k most similar observations from the training data to predict the category or continuous outcome of observations in the test set. This process is repeated many times with randomly re-selected training and test sets to cross-validate its findings.

Dividing the data into training and test sets is done by randomly allocating some percentage of the data into each group. In this case, I assign 70 percent of the data to the training set and 30 percent to the test set. When this sampling is repeated during cross-validation, it is done with replacement. In order to calculate the ideal k for each training set, I use the `cv.train` function in the `kknn` package in R (Schliep and Hechenbichler, 2014). Finally, I use the `kknn` function in that package to predict each test set school's average test score using the previously established k and the Euclidean distances between observations. In order to prevent variables with large values from overweighting those with smaller values, I preceded this analysis by standardizing all features of all observations such that they range between zero and one (Tanner and Toivonen, 2010). This is accomplished by z-score standardization, or by dividing the difference between the observed variable and the maximum for that variable into the difference between the variable's minimum and maximum values. Finally, I performed this standardization for test scores within school type, as each school type (primary, combined, secondary) is expected to score differently on their different NAPLAN assessments.

In an exploratory element to this analysis, I also create an additional funding column in which I add 500 Australian dollars to each school's per-student funding. This is done before standardization, which I perform with the original minimum value in order to prevent the change from disappearing. In every iteration, I train KNN on the training set and test it by predicting the scores of schools in the test set. Following this, I predict scores for precisely the same group of test schools with every variable the same except for increased funding. I predict how the schools would perform with hypothetical extra funding based on how real schools actually perform. By doing this, I can compare the predicted score of each test school to its predicted score with additional funding and look for patterns.

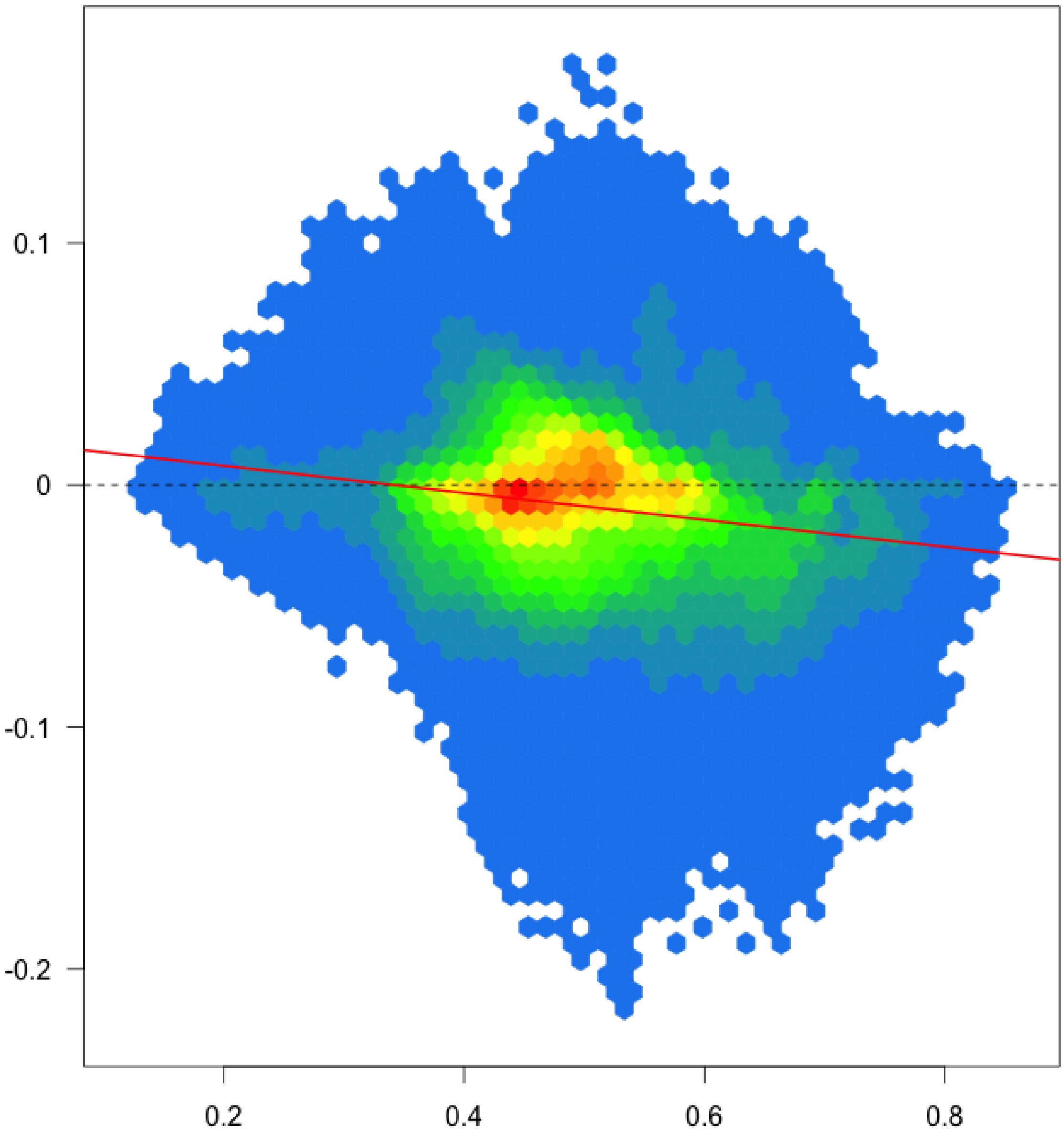
The biggest potential downfall of KNN and statistical learning techniques in general is the risk of overfitting predictions to the training data, so it is critically important to cross-validate any decisions imposed on the method (James et al., 2013). For this reason, I cross-validate the analysis by resampling test and training sets with replacement 500 times. Because I combine three years of data and sample unevenly between the training and test sets, it is highly likely that a given school will be a near neighbor to itself at a different time period. This adds additional strength to the exploratory part of the analysis as predictions are based on similar schools as well as potentially the same school in different circumstances. The end result is a data frame with more than 800,000 observations, each of which includes a predicted test score for the school with its real funding and with hypothetical increased funding.

KNN has been used before in educational contexts, both in its classification implementation and its regression implementation, in which the actual grade or score is predicted instead of only passing and failing. KNN is especially useful for education because it functions well when data is noisy, incomplete, and otherwise complex (Zou and Huang, 2005). For example, Kotsiantis and Pintelas (2003) use KNN to classify students and predict dropout. Minaei-Bidgoli et al. (2003) use KNN to predict student scores in an online course. Finally, Shih and Lee (2001) attempt to develop an adaptive learning system using the predictions from their KNN analysis. This paper builds on their work by raising the unit of analysis to the school level and using the method as a tool for understanding how predictions in school achievement outcomes change with increased funding.

4.5 Results

When the predictions of school achievement with normal funding and additional funding are compared, the results indicate that a large funding increase of 500 Australian dollars per student changes predicted scores differently across the predicted achievement distribution. For schools at the highest and lowest predicted achievement levels, the funding increase is associated with very small changes in predicted school performance—less than 5 percent change in the top and bottom quartiles. In contrast, the schools in the middle of the predicted achievement distribution are greatly varied in their predicted response to funding changes, and might see changes as large as 10-15 percent improvement or 10-20 percent contraction in the middle half of the predicted achievement distribution. Score predictions for schools at the highest and lowest ends of the predicted achievement distribution hardly react at all to increased funding, while those at in the middle can change but with great variability—see Figure 4.1. Figures are generated using the `hexbin` package in R (Carr and Maechler, 2013).

Figure 4.1: Prediction Change by Predicted Achievement



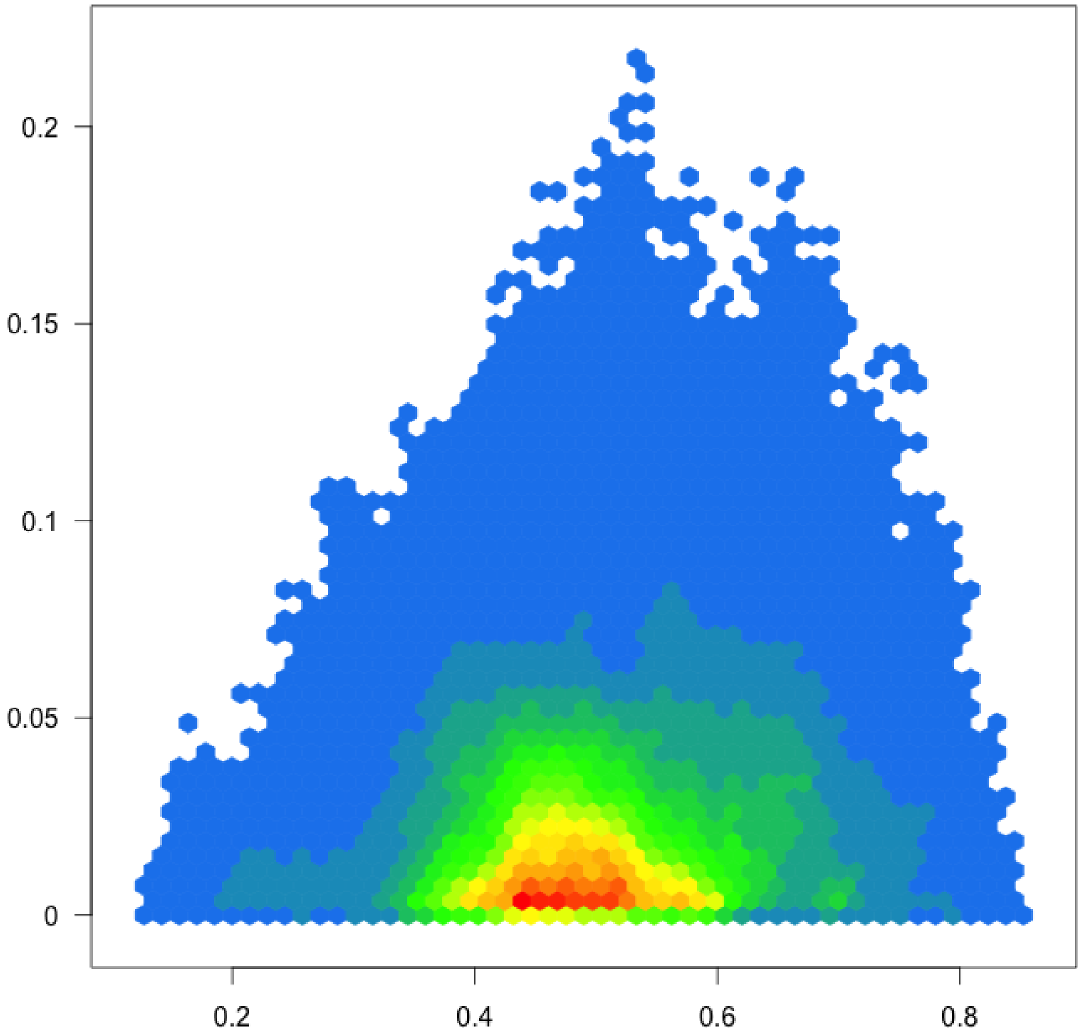
In general, changes in predicted score from increased funding are small—especially for the worst and best schools. Improvement is relatively symmetric along the predicted achievement distribution with schools closer to the middle seeing more potential change and those at the ends seeing less. For cases when predicted scores get worse, however, there is asymmetry and the schools around the median and in the third quartile of predicted achievement seem to drive a negative overall trend in predicted score changes as real predicted scores improve. The simple linear regression line reflects this, starting slightly above zero at a 1.9 percent improvement for the lowest-performing schools and sloping downward with a slope of -0.056 until there is a negative change for the schools with the highest predicted scores. The schools at the lower end of the predicted test score distribution might expect to see slightly improved outcomes on average, while those at the higher end could get worse. Although this regression is extremely significant, it is merely an indication of the general trend in the data and should not be considered as a causal analysis. Details of that regression are in Table 4.2.

Table 4.2: Prediction Change by Predicted Achievement

	Dependent variable:
	diff
Change in Predicted Achievement	-0.056^{***}
	(0.0003)
Constant	0.019^{***}
	(0.0002)
Observations	807,500
R^2	0.034
Adjusted R^2	0.034
Residual Std. Error	0.031 (df = 807498)
F Statistic	$28{,}508.580^{***}$ (df = 1; 807498)
Note:	*p<0.1; **p<0.05; ***p<0.01

Differences in effects sizes along the achievement distribution are clearest in a plot of root squared change between predictions based on true funding data and predictions based on data with increased funding. The plot shows a clear inverted-U shaped pattern of change, with the schools at the low and high ends of the predicted achievement distribution hardly changing their predicted scores while predicted scores in the middle move much more dramatically and with great variability. Figure 4.2 shows this pattern.

Figure 4.2: Absolute Prediction Change by Predicted Achievement



In general, a hypothetical increase in funding is not associated with improved outcomes across the board as would have been expected. Increased funding does not, it seems, always lead to improved outcomes. Instead, its effects on predictions are very heterogeneous, especially in the center of the achievement distribution. These effects are also quite small considering the size of the funding increase, and are smallest for the schools with the highest and lowest predicted scores.

4.5.1 Long-Term Changes

School funding changes can take time to turn into actual changes for students and teachers. The ACARA data reflects this by using funding data from the previous year in each year's school profile data. I extend that concept by using test scores from 2012 as the outcome variable with school profile and financial data from 2010. This limits me to the use of only one year's data, but the smaller data size is still adequate for this analysis and the results will be useful for identifying how the pattern of change in predicted test scores persists or changes over time. I repeat the same analytical process, this time emerging with 333,000 schools with predictions for test scores in 2012 given their real funding in 2010 and that funding plus 500 Australian dollars. For summary statistics of this dataset, see Table C.6 in the Appendix.

As was the case with the analysis of one-year effects, this longer-term analysis shows a pattern of very minimal changes in achievement predictions at the lowest and highest ends of the predicted achievement distribution with larger and more variable change in the middle (see Figure 4.3). In the longer term, however, the change in predicted achievement is more positive on average and especially for the lowest-achieving schools (see Table 4.3). The intercept here is a three percent improvement for the worst-performing schools (up from 1.9 percent in the one-year predictions) and the slope is less negative at -0.05 instead of -0.056. Again, the pattern of prediction change is fairly symmetric for school improvement and asymmetric for contracting schools as the schools in the third quartile have larger contractions than those in the second. However, this time there is a market improvement among some schools in the lowest quartile—an improvement that survived cross-validation so must be considered significant. Similarly, there appear to be fewer struggling schools in the third quartile than there were before and that pattern is less marked. Over the long term, schools' achievement prediction changes are more positive than they are in the short term, and there is more evidence that the lowest-performing schools are improving.

Figure 4.3: Two-Year Prediction Change by Predicted Achievement

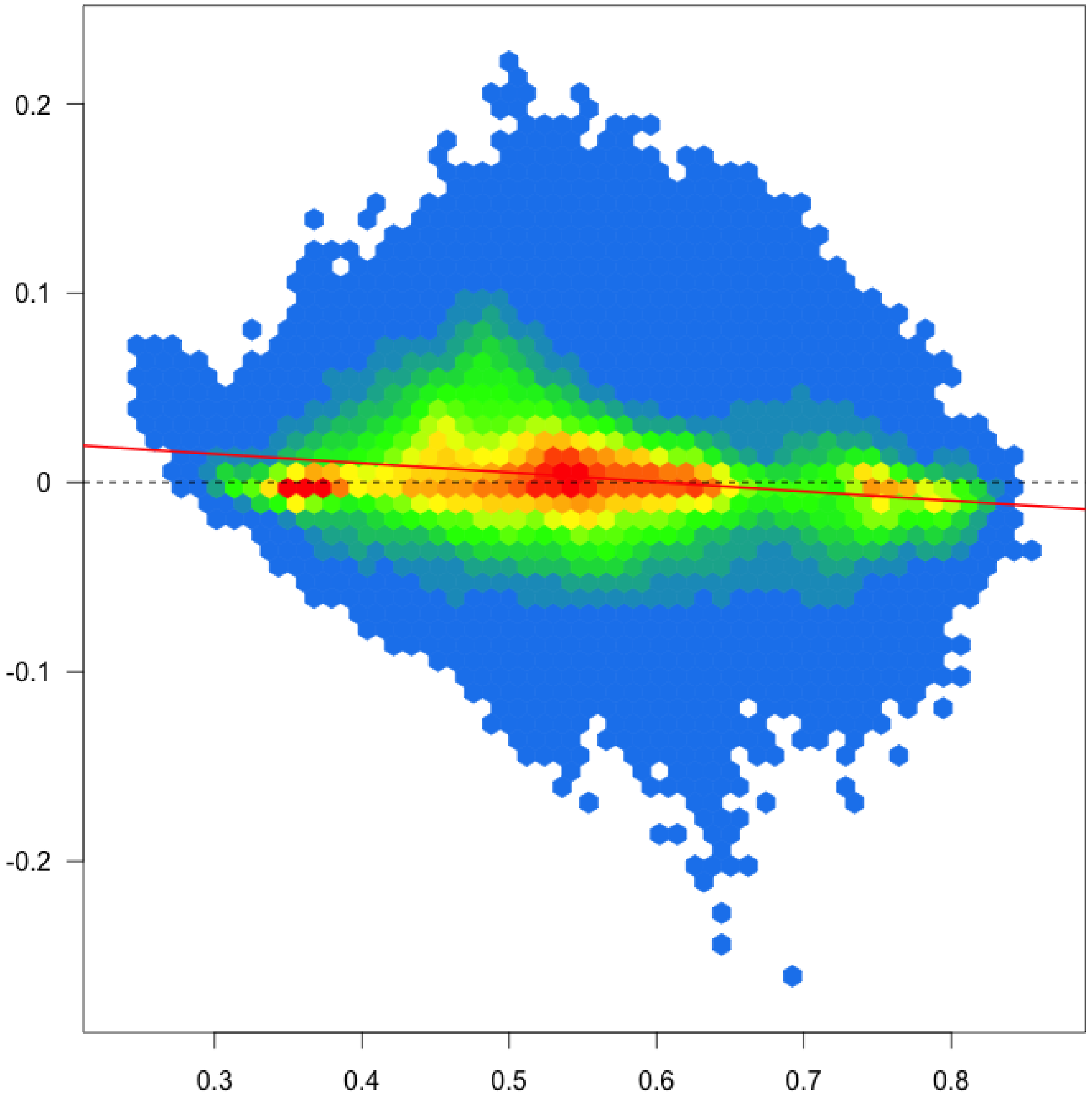
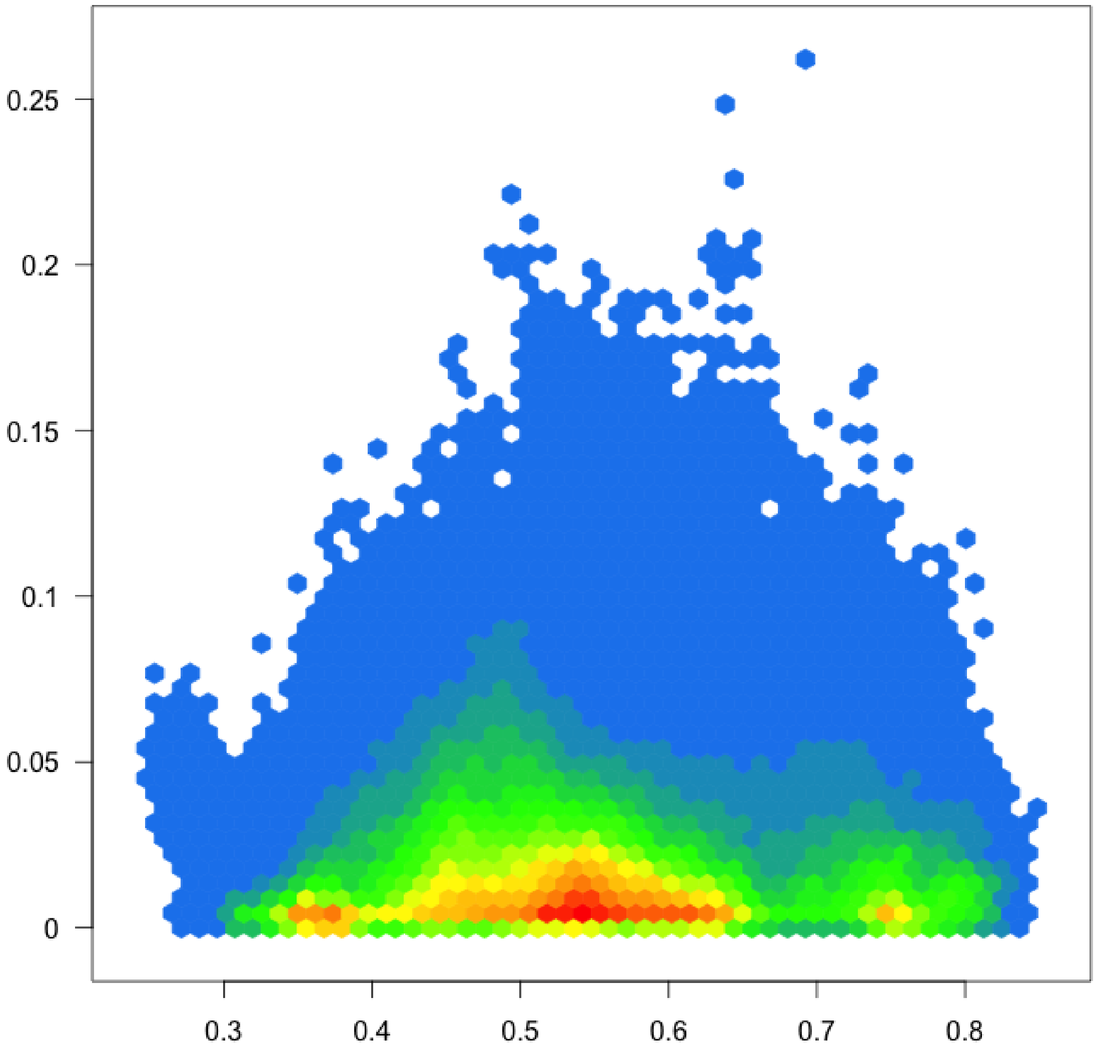


Table 4.3: Two-Year Prediction Change by Predicted Achievement

	Dependent variable:
	diff
Change in Predicted Achievement	-0.050^{***}
	(0.0004)
Constant	0.030^{***}
	(0.0002)
Observations	333,000
R^2	0.039
Adjusted R^2	0.039
Residual Std. Error	0.029 (df = 332998)
F Statistic	$13,387.300^{***}$ (df = 1; 332998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Finally, the plot of absolute change in predicted test scores by predicted achievement level is again inverse-U-shaped, but this time it is fatter—there is more change among all schools rather than only the middle schools (see Figure 4.4). The pattern of unusual improvement in the lowest-performing schools is also evident on this plot. Over the long term, the general pattern holds true but predicted change is generally larger at the tails and more positive, especially for the lowest-performing schools.

Figure 4.4: Two-Year Absolute Prediction Change by Predicted Achievement



4.6 Discussion

The results discussed previously have useful implications for our understanding of educational production in the context of school finance over the short and long term for educational policy makers interested in using the data they collect to help inform policy decisions. As discussed earlier, economic theory predicts that increasing inputs should improve educational quality. I now consider what could be driving the short- and long-term results and discuss how this analysis is relevant for educational policy makers in general in their use of school data.

The general shape of the change in predicted outcomes along the predicted achievement distribution as funding is increased is fairly intuitive and consistent with economic theory. While theory might predict that educational quality will improve with greater quantities of funding, it would also predict that structural limitations and diminishing marginal returns would hamper the improvement of the worst and best schools, respectively. It is possible to imagine scenarios in which the worst and the best schools invest additional funding efficiently and still fail to see improved test scores. At the same time, there are scenarios where inefficient investment could be the root of the problem. The variation in predicted school performance changes among schools in the middle of the distribution presents its own set of potential scenarios. Each presents its own set of appropriate policy responses and challenges.

The schools with the lowest predicted performance are likely to share other issues beyond their test scores, like a lack of the basic infrastructure for efficiently educating children. If additional funding is used to hire competent staff and teachers, to fix major issues with school facilities, or to solve other basic but not instructional issues, it is possible that this allocation of funds is—while indisputably correct—not associated with improvement in achievement levels in the year immediately following added funding. Of course it is also possible that some of these schools may be poorly managed and might fail to use resources efficiently if the school council, principal, and other school leadership are doing their jobs poorly. The spike in improvement among the very worst schools over the long term suggests that these schools' investments in infrastructure and other necessities does indeed raise test scores, but only once the construction, hiring, and other restructuring projects are complete.

This result is very optimistic for the lowest-achieving schools despite the apparent indication that the worst schools do not improve with additional resources—they simply need more time and perhaps more guidance. These results, therefore, do not suggest that money spent on low-achieving schools is a wasted investment, but rather that the schools should receive something more than simply more resources. A longer timeframe for improvement expectations and some means of measuring structural needs and school management competence may help turn financial investment in struggling schools into measurable improvement in educational quality.

Schools at the top of the predicted achievement distribution also fail to show significant improvement, but may do so due to the diminishing marginal returns on resources. These schools are already high performers, and improving their performance by even 5 percent would simply require a massive investment of resources, good management, quality instruction, and effort on the part of all school leadership, staff, teachers, students, and parents—a much larger effort than a lower-performing school might require. Given that the schools are already performing more than adequately well, a small and expensive improvement like this is not necessarily efficient. This is absolutely not to say that money should not be spent on the best schools—these schools likely need additional resources simply to maintain their already-high levels of achievement—and again should not be taken as an indication that schools at the top end do not need more resources. They may need help just to stay where they are—evident in the slight negative change in their predicted performance on average. Furthermore, the negative change is still present (though smaller) in the long-term specification, so school success at a given time is no guarantee of later high scores. Some schools that perform well for a short time will regress to the mean, and the rapidly changing nature of student bodies emphasizes that their high performance should not be taken for granted.

The schools in the middle of the predicted achievement distribution are possibly the most interesting part of these findings. Their absolute level of change is much higher than the schools on the bottom and top, implying that financial investment is a policy tool that can produce the most change among mid-level schools. However, that change is not always positive. The pattern shown by the schools whose predicted scores change positively is predictable by economic theory: it is an inverse-U-shaped curve that reflects both structural issues at the bottom end and diminishing returns at the top. The middle schools—who are far enough from the bottom to avoid structural issues and far enough from the top to avoid suffering from diminishing returns—show the greatest change. However, the picture is much less predictable for schools whose predicted scores change negatively with increased funding. This pattern is skewed such that the mass of the graph is in the third quartile; schools that are doing relatively well between the median and the top quartile have the widest negative range.

The issue of what might be driving schools who get worse with increased funding is one that is not easily predicted by economic theory but is quickly understood by education scholars and educators themselves. Money is not necessarily an instructional resource, and school spending is not limited to the activities of teaching and learning. Schools in the second quartile—between the bottom and the median—might be struggling enough that they face sanctions or parent pressure and must focus on raising test scores, or that there remain obvious holes in their ability to practice teaching and learning. In these cases, the school is either strongly incentivized to funnel resources into achievement or its administrators can easily detect opportunities to invest their new resources into teaching and learning directly. In some cases, these changes may be done poorly or the school may face new challenges—there are still schools in the second quartile that contract with increased funding, but they do so less dramatically than the third quartile.

Schools in the third quartile have the greatest variability in their predicted responses to increased funding, driven by the much more negative lower bound for school contraction among these schools than others. The schools in the third quartile might not face such strong test score-related pressure or have such obvious holes in their teaching and learning activities, and may use extra money for projects that either fail to improve the activities of teaching and learning or actively detract from them. For example, a school might invest in a new theater or gym that enhances the students' lifestyles but does not improve test scores, or it might buy new technology for the classroom that actually takes away from instructional time as teachers and students struggle to operate their new tools efficiently. These stories are supported by the long-term findings, as the schools in the third quartile are have less dramatic negative changes two years later—the construction project is over and the teachers have learned to incorporate their new tools in the classroom. If policy makers want to minimize the potential that additional financial investments have the opposite of their intended effects, it may be important to either earmark additional funding for teaching and learning specifically, or to provide strong positive achievement incentives for schools that perform above the minimum standards. Of course, educational quality is not always best measured by test scores, but that conversation is beyond the scope of this paper.

4.7 Conclusions

This study suggests that school performance does respond in relatively predictable ways to increased funding, but also underscores the complexity of predicting educational outcomes. Schools do improve in a way that matches the predictions of economic theory when given additional funding, so long as the limitations of inadequate structures and diminishing returns are taken into account. However, schools performance might also contract when additional funding is granted. This is not as easily predicted by purely economic models, but can be understood in the concept of education scholarship: poor school management and use of funds for projects that do not improve or even detract from teaching and learning activities can contribute to schools performing worse when given extra money. In economic terms, this is inefficient investment of resources. Policy makers can take this pattern into account when determining how best to improve educational quality.

On a further policy note, this suggests a means by which educational policy makers can use register data collected by the school system to inform educational policy making.

The method used here is simple to understand and is model-free, so it could be easily constructed and interpreted by a policy maker. With so much emphasis on data collection and so many resources dedicated to collecting, tabulating, and holding data, it is increasingly important to find ways to effectively use that data to improve educational quality. This paper demonstrates one way that a simple method associated with big data can be used to illuminate theory while informing policy.

This analysis has limitations. It is not a causal analysis, only an exploration of how schools' achievement predictions change given additional funding. The algorithm has the advantage of being model-free and therefore free from the assumptions of researchers, but that is also a disadvantage because it makes the inclusion of theory more difficult. For example, all school features are equally weighted in this analysis to avoid imposing the assumptions of previous analyses, but this might reduce the accuracy of predictions. In addition, this analysis is only a comparison between predicted scores and adjusted predictions—it would need to be tested on real administration decisions before its utility can be conclusively stated.

This type of analysis can be performed by school administrators and policy makers with free, simple tools and the data they have on hand. Using an analysis like this one, educational officials may be able to identify areas where school intervention through increased funding is most effective and where additional incentives and policies would be required. Schools at the lowest achievement levels, for example, may require infrastructure or leadership changes before simple financial investments can be used effectively. Those in the second quartile can make changes, but might require some guidance in their use of additional resources. Schools in the third quartile can improve greatly but can also struggle to invest additional resources well, so maintaining a focus on teaching and learning while providing additional resources is important. Finally, schools in the top quartile of the achievement spectrum are effectively using what they have and may have more specific needs for individual programs, or they might need help to keep from backsliding. Real-world school administrators can use analyses like this to help efficiently allocate funding for achievement improvement.

Chapter 5

Final Remarks

Educational policy makers face a complex universe of choices when attempting to improve educational outcomes like achievement and equity. By utilizing the information available to them in creative ways, this complexity can be greatly reduced. Approaching educational data using a variety of methodological, strategic, and temporal tools can help to illuminate problems from multiple perspectives and in new ways. I have described here three research studies that each take a different perspective on research method, purpose, and time in order to illustrate this point. All three are empirical, data-driven, and produce results that can be used by policy makers to justify specific policy choices. Rigor and empiricism in the context of education economics are not limited to a single method, a single set of research questions, or a single timeframe; the three studies detailed here demonstrate that point.

The three studies are diverse: each one uses a different method to evaluate a separate key issue while looking in distinct temporal directions. "Is funding enough? A configurational analysis of conditions for school achievement in Victoria, Australia," uses QCA to evaluate educational equity in the present. "The impact of high school exit exams on graduation rates and achievement" uses parametric and non-parametric interrupted time series specifications to conduct policy analysis over a long period of time in the past. Finally, "Predicting school achievement reactions to increased funding" uses KNN—a technique most associated with big data analytics—to investigate the role of education finance in the future performance of schools. In turn, each of these studies reflects its own segment of the methods, key issues, and temporal perspectives that can be used by economists of education in their work.

The three methods include a set-theoretic strategy, a regression-based econometric analysis, and a model-free data mining tool. Education economists are not limited to a single family of methods, and each of these is useful for answering certain types of questions from data with specific characteristics. No single method is sufficient by itself for answering all of the questions important to education economists, and each of these illustrates the utility of its family of methods.

In addition to utilizing methods from three epistemological foundations, I have addressed the three broad issues relevant to education economics: equity, policy analysis, and efficiency or educational production. Equity is the main topic of the first study; dealt with in this case by examining the necessary and sufficient conditions for school success and failure. The second study looks specifically at one high-level educational policy, upon whose implications sociological and economic theories disagree. This demonstrates the utility of economics as a framework for understanding educational policy while examining the real-world impact of the policy in question. Finally, the third study addresses educational production by predicting school-level achievement changes in the case of added funding. This is also part of the broader conversation on efficiency in education, as researchers and policy makers attempt to create the highest possible educational quality with the resources available. Though equity, policy analysis, and educational production are all complex topics that encompass myriad research questions, they are the key issues in education economics and are each addressed by the three studies presented here.

Education research cannot only focus on one period in time, as the current, past, and potential performance of an education system are all important. Examining current performance with an eye to the impact of short-term adjustments is useful and must be practiced by educational authorities. However, this narrow focus could potentially lead to the broader trends being buried in the granular data. At the same time, using longitudinal data to look at past policies can tell us a great deal about the performance of policies over time, but risks missing the factors at work on children currently in school. Prediction of school performance in the future is often ignored, but can be a useful means of identifying mechanisms and next steps. Still, it must be grounded in an understanding of the policies that have worked in the past and how they affect current schools and schooling.

Education research often refers to its own complexity—for good reason—but rigorous analysis of the great deal of available data can help to untangle at least some of the knots. This dissertation explores three methodologies available to researchers, three key topics for evaluation, and three ways to look at educational policy questions. All of these studies answer their own research questions, and they all lead to further inquiries and future research as well. How do the profiles of successful and failing schools change over time? Will the application of new HSEE policies generate different graduation rate trends, or follow the trends we describe? How might the predicted reactions of school achievement to additional funding align with an actual funding increase? From these, even broader questions can be derived—questions that concern the entire fields of education and education economics. How can we use educational policy to improve equity? What are the real short- and long-term effects of educational policies? And what are the real parameters that govern the relationship between educational resources and achievement?

Education is undoubtedly a complicated issue, and researchers and policy makers can still draw actionable insights by creatively and thoughtfully utilizing educational data. Ever more data is available, as are increasingly sophisticated tools for its analysis. By combining these resources, policy makers and scholars can understand and drive educational equity, policy, and efficiency.

Bibliography

ACARA. Guide to understanding icsea. Technical report, ACARA Copyright Administration, Australia, 2012.

G. A. Akerlof. The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, August 1970. URL http://ideas.repec.org/a/tpr/qjecon/v84y1970i3p488-500.html.

A. L. Amrein and D. C. Berliner. High-stakes testing & student learning. *Education Policy Analysis Archives*, 10:1–74, 2002.

P. Arcidiacono, P. Bayer, and A. Hizmo. Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4): 76–104, 2010.

S. Balestra and U. Backes-Gellner. When a door closes a window opens? Investigating the effects of involuntary separations. Economics of Education Working Paper 0072, University of Zurich, Institute for Strategy and Business Economics (ISU), February 2012. URL http://ideas.repec.org/p/iso/educat/0072.html.

S. Balestra and U. Backes-Gellner. Revisiting class-size effects: Where they come from and how long they last. *Economics of Education Working Paper Series*, 102, 2014.

B Bandaranayake. Formula-based school funding system in victoria: An empirical analysis of equity. *NCPEA International Journal of Educational Leadership Preparation*, 8 (2):191–207, 2013.

J. H. Bishop. The effect of national standards and curriculum-based exams on achievement. *American Economic Review*, 80(2):260–264, 1997.

J. H. Bishop and F. Mane. The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success. *Labour Economics*, 8(2):203–222, May 2001. URL `http://ideas.repec.org/a/eee/labeco/v8y2001i2p203-222.html`.

J. H. Bishop, F. Mane, and M. Bishop. How external exit exams spur achievement. *Educational Leadership*, 59(1):58–63, 2001.

M. P. Bitler, J. B. Gelbach, and H. W. Hoynes. What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, 96(4):988–1012, September 2006. URL `http://ideas.repec.org/a/aea/aecrev/v96y2006i4p988-1012.html`.

H. S. Bloom, S. Ham, L. Melton, and J. O'Brien. Evaluating the accelerated schools approach: A look at early implementation and impacts on student achievement in eight elementary schools. Evaluative Report 142, Manpower Demonstration Research Corporation, MDRC, 2001.

J. Booher-Jennings. Below the bubble: "educational triage" and the texas accountability system. *American Educational Research Journal*, 42(2):231–268, 2005.

M. Carnoy and S. Loeb. Does external accountability affect student outcomes? a cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4):305–331, 2002.

N. Carr, D. Lewin-Koh and M. Maechler. *hexbin: Hexagonal Binning Routines*, 2013. URL `http://CRAN.R-project.org/package=hexbin`. R package version 1.26.3.

B. Cooper. Applying ragins crisp and fuzzy set qca to large datasets: Social class and educational achievement in the national child development study. *Sociological Research Online*, 10(2), 2005.

B. Cooper and J. Glaesser. How has educational expansion changed the necessary and sufficient conditions for achieving professional, managerial and technical class positions in britain? a configurational analysis. *Sociological Research Online*, 13(3), 2008.

B. Cooper and J. Glaesser. Contrasting variable-analytic and case-based approaches to the analysis of survey datasets: Exploring how achievement varies by ability across configurations of social class and sex. *Methodological Innovations Online*, 5(1):4–23, 2010.

B. Cooper and J. Glaesser. Paradoxes and pitfalls in using fuzzy set qca: Illustrations from a critical review of a study of educational inequality. *Sociological Research Online*, 16(3), 2011.

T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

T. S. Dee and B. Jacob. Do high school exit exams influence educational attainment or labor market performance? NBER Working Paper 12199, National Bureau of Economic Research, Inc, May 2006. URL `http://ideas.repec.org/p/nbr/nberwo/12199.html`.

T. S. Dee and B. Jacob. The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3):418–446, Summer 2011. URL `http://ideas.repec.org/a/wly/jpamgt/v30y2011i3p418-446.html`.

T. S. Dee, W. N. Evans, and S. E. Murray. Data watch: Research data in the economics of education. *Journal of Economic Perspectives*, 13(3):205–216, Summer 1999. URL `http://ideas.repec.org/a/aea/jecper/v13y1999i3p205-216.html`.

T. S. Dee, B. Jacob, and N. L Schwartz. The effects of nclb on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2):252–279, 2012.

DEECD. Financial manual for victorian government schools. Technical report, Department of Education and Early Childhood Development, Melbourne, Australia, 2011.

DEECD. The student resource package 2013 guide. Technical report, Department of Education and Early Childhood Development, Melbourne, Australia, 2012.

J. Dewey, T. A. Hosted, and L. W. Kenny. The ineffectiveness of school inputs: A product of misspecification? *Economics of Education Review*, 19(1):27–45, 2000.

S. Doric. Ideas and education: Level or growth effects? *National Bureau of Economic Research*, page NBER Working Paper No. 9709, 2003.

V. Frisancho, K. Krishna, S. Lychagin, and C. Yavas. Better luck next time: Learning through retaking. IDB Working Paper 483, Inter-American Development Bank, 2013. URL `http://ideas.repec.org/p/nbr/nberwo/12199.html`.

R. Gibbons and L. F. Katz. Layoffs and lemons. *Journal of Labor Economics*, 9(4):351–380, October 1991. URL `http://ideas.repec.org/a/ucp/jlabec/v9y1991i4p351-80.html`.

J. Glaesser and B. Cooper. Selectivity and flexibility in the german secondary school system: A configurational analysis of recent data from the german socio-economic panel. *European sociological review*, 27(5):570–585, 2011.

J. Glaesser and B. Cooper. Exploring the consequences of a recalibration of causal conditions when assessing sufficiency with fuzzy set qca. *International Journal of Social Research Methodology*, 17(ahead-of-print):1–15, 2013.

D. Gonski, K. Boston, K. Greiner, C. Lawrence, B. Scales, and P. Tannock. Review of funding for schooling—final report. Technical report, Commonwealth Government, Canberra, Australia, Canberra, Australia, December 2011.

C W J Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, July 1969. URL `http://ideas.repec.org/a/ecm/emetrp/v37y1969i3p424-38.html`.

C. W. J. Granger. Some recent development in a concept of causality. *Journal* *of Econometrics*, 39(1-2):199–211, 1988. URL `http://ideas.repec.org/a/eee/econom/v39y1988i1-2p199-211.html`.

J. P. Greene and M. A. Winters. Public high school graduation and college-readiness rates: 1991-2002. Education Working Paper 8, Center for Civic Innovation, CCI, 2005.

E. Grodsky, J. R. Warren, and D. Kalogrides. State high school exit examinations and naep long-term trends in reading and mathematics, 1971-2004. *Educational Policy*, 23 (4):589–614, 2009.

E. A. Hanushek. The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, 24(3):1141–1177, 1986.

E. A. Hanushek. *Handbook of the Economics of Education*, volume 2, chapter Chapter 14 School resources, pages 865–908. Elsevier, 2006.

E. A. Hanushek, S. J. Machin, and L. Woessmann, editors. *Handbook of the Economics of Education*, volume 4. Elsevier, 2011.

E. Jackson and M. E. Page. Estimating the distributional effects of education reforms: A look at Project STAR. *Economics of Education Review*, 32(C):92–103, 2013. URL `http://ideas.repec.org/a/eee/ecoedu/v32y2013icp92-103.html`.

B. A Jacob. Getting tough? the impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2):99–121, 2001.

L. S. Jacobson, R. J. LaLonde, and D. G. Sullivan. Earnings losses of displaced workers. *American Economic Review*, 83(4):685–709, September 1993. URL `http://ideas.repec.org/a/aea/aecrev/v83y1993i4p685-709.html`.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning.* New York: Springer, 2013.

B. Jensen and J. Sonnemann. Turning around schools: It can be done. Technical report, Grattan Institute, Melbourne, Australia, 2014.

D. Johnson, B. S. Jensen, and B. Methakullawat. Multivariate analysis of performance of victorian schools. In *Melbourne, 26 August*. Making Schools Better Summit Conference, 2004.

C. J. Kotsiantis, S. B. Pierrakeas and P. E. Pintelas. *Knowledge-Based Intelligent Information and Engineering Systems*, chapter Preventing student dropout in distance learning using machine learning techniques, pages 267–274. Springer Berlin Heidelberg, 2003.

A. B. Krueger. Economic considerations and class size. *The Economic Journal*, 113: F34–F63, 2003.

S. Lamb, A. Walstab, R. Teese, M. Vickers, and R. Rumberger. Staying on at school: Improving student retention in australia. Technical report, Queensland Department of Education the Arts, Queensland, Australia, 2004.

R. Levacic. *Governance and performance of education systems*. Springer Netherlands, 2007.

R. Levacic and P. Downes. Formula funding of schools, decentralization and corruption: A comparative analysis. Technical report, International Institute for Educational Planning (IIEP) UNESCO, Paris, France, 2004.

R. Levacic, K. Ross, B. Caldwell, and A. Odden. Funding schools by formula: Comparing practice in five countries. *Journal of Education Finance*, pages 489–515, 2000.

G. J. Marchant and S. E. Paulson. The relationship of high school graduation exams to graduation rates and sat scores. *Education Policy Analysis Archives*, 13(6):1–17, 2005.

G. N. Marks. What aspects of schooling are important? school effects on tertiary entrance performance. *School Effectiveness and School Improvement*, 21(3):267–287, 2010.

P. Miller, C. Mulvey, and N. Martin. What do twins studies reveal about the economic returns to education? a comparison of australian and us findings. *American Economic Review*, 85(3):586–599, 1995.

P. W. Miller and D. Voon. Lessons from my school. *Australian Economic Review*, 44(4): 366–386, 2011.

B. Minaei-Bidgoli, D. A. Kashy, G. Kortmeyer, and W. F. Punch. Predicting student performance: An application of data mining methods with an educational web-based system. In *Frontiers in Education Conference, 33rd Annual*, volume 1, pages T2A–13. IEEE, 2003.

R. J. Murnane. U.S. high schoolig graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51(2):370–422, June 2013. URL http://ideas.repec.org/a/aea/jeclit/v51y2013i2p370-422.html.

OECD. Education at a Glance 2013: OECD Indicators. Technical report, OECD, Paris, France, 2013. URL http://dx.doi.org/10.187/eag-2013-en.

D. Ou. To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam. *Economics of Education Review*, 29(2):171–186, April 2010. URL http://ideas.repec.org/a/eee/ecoedu/v29y2010i2p171-186.html.

A. D. Plunk, W. F. Tate, L. J. Bierut, and R. A. Grucza. Intended and unintended effects of state-mandated high school science and mathematics course graduation requirements on educational attainment. *Educational Researcher*, 43(5):230–241, 2014.

K. Poder, K. Kerem, and T. Lauri. Efficiency and equity within european education systems and school choice policy: Bridging qualitative and quantitative approaches. *Journal of School Choice*, 7(1):1–36, 2013.

A. Preston. Where are we now with human capital theory in australia? *Economic Record*, 73(220):51–78, 1997.

K. Pugh and G. Foster. Australia's national school data and the "big data" revolution in education economics. *Australian Economic Review*, 47(2):258–268, 2014.

C. C. Ragin. *The comparative method: Moving beyond qualitative and quantitative strategies*. Univ of California Press, 1987.

C. C. Ragin. *Fuzzy-set social science.* Chicago: University of Chicago Press, 2000.

C. C. Ragin. *Redesigning social inquiry: Fuzzy sets and beyond.* Chicago: University of Chicago Press, 2008.

S. F. Reardon, N. Arshan, A. Atteberry, and M. Kurlaender. Effects of failing a high school exit exam on course taking, achievement, persistence, and graduation. *Educational Evaluation and Policy Analysis*, 32(4):498–520, 2010.

C. Romero and S. Ventura. Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6):601–618, 2010.

A. Schenker-Wicki. *Governance and performance of education systems*, chapter Comment: Funding schools by formula, Levacic, R. Springer, 2008.

K. Schliep and K. Hechenbichler. *kknn: Weighted k-Nearest Neighbors*, 2014. URL `http://CRAN.R-project.org/package=kknn`. R package version 1.2-5.

C. Q. Schneider and C. Wagemann. *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis.* Cambridge University Press, 2012.

W. R. Shadish, T. D. Cook, and D. T. Campbell. *Experimental and quasi-experimental designs for generalized causal inference.* Wadsworth Cengage Learning, 2002.

B. Y. Shih and W. I. Lee. The application of nearest neighbor algorithm on creating an adaptive on-line learning system. In *Frontiers in Education Conference, 31st Annual*, volume 1, pages T3F–10. IEEE, 2001.

B. W. Silverman and M. C. Jones. E. fix and jl hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951). *International Statistical Review/Revue Internationale de Statistique*, 1951:233–238, 1989.

N. C. Soguel and P. Jaccard, editors. *Governance and performance of education systems*, volume 1. Springer London, Limited, 2008.

A. M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, August 1973. URL `http://ideas.repec.org/a/tpr/qjecon/v87y1973i3p355-74.html`.

T. Tanner and H. Toivonen. Predicting and preventing student failure—using the k-nearest neighbour method to predict student performance in an online course environment. *International Journal of Learning Technology*, 5(4):356–377, 2010.

A. Thiem and A. Dusa. Qca: A package for qualitative comparative analysis. *The R Journal*, 5(1):87–97, 2013.

P. E. Todd and K. I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485):F3–F33, 2003.

L Tolstoy. *Anna Karenina*. Harmondsworth, 1954, 1877.

T. M. Trujillo and S. L. Woulfin. Equity-oriented reform amid standards-based accountability a qualitative comparative analysis of an intermediary?s instructional practices. *American Educational Research Journal*, 51(2):253–293, 2014.

J. H. Tyler, R. J. Murnane, and J. B. Willet. Estimating the labor market signaling value of the GED. *Quarterly Journal of Economics*, 115(2):431–468, May 2000. URL `http://www.jstor.org/stable/2586999`.

T. von Wachter, J. Song, and J. Manchester. Long-term earnings losses due to job separation during the 1982 recession: An analysis using longitudinal administrative data from 1974 to 2004. Discussion Paper 0708-16, Columbia University, Department of Economics, New York, 2008. URL `http://ideas.repec.org/p/clu/wpaper/0708-16.html`.

J. R. Warren and M. R. Edwards. High school exit examinations and high school completion: Evidence from the early 1990s. *Educational Evaluation and Policy Analysis*, 27(1):53–74, 2005.

J. R. Warren, K. N. Jenkins, and R. B. Kulick. High school exit examinations and state-level completion and ged rates, 1975 through 2002. *Educational Evaluation and Policy Analysis*, 28(2):131–152, 2006.

A. Zou, Y. An and X. Huang. Evaluation and automatic selection of methods for handling missing data. In *IEEE International Conference on Granular Computing*, volume 2, pages 728–733. IEEE, 2005.

Appendix A

APPENDIX

A.1 QCA Truth Tables

Table A.1: Truth Table for Positive Outcome Analysis

ICSEA	URBAN	SIZE	TEACH	ATTEND	FUND	OUT	n	Incl.	PRI
1	1	1	0	1	1	1	148	0.921	0.895
0	0	0	1	0	0	0	82	0.412	0.234
0	1	0	1	0	0	0	29	0.412	0.124
0	1	1	1	0	0	0	29	0.430	0.137
0	0	0	1	1	0	0	19	0.564	0.309
1	0	0	1	0	0	0	19	0.612	0.418
1	1	0	0	1	1	1	18	0.900	0.844
0	0	0	1	0	1	0	17	0.483	0.250
1	1	1	1	1	1	1	17	0.904	0.846
0	1	1	0	0	0	0	15	0.478	0.165
1	0	0	1	1	0	0	14	0.700	0.482
0	0	0	1	1	1	0	13	0.608	0.326
1	0	0	1	0	1	0	13	0.604	0.389
1	1	0	1	1	1	1	13	0.872	0.791
0	0	0	0	0	0	0	12	0.457	0.198
1	0	0	1	1	1	0	12	0.700	0.486
1	0	1	0	1	1	0	12	0.734	0.459
1	1	1	0	1	0	0	12	0.800	0.651
0	0	1	1	0	0	0	11	0.395	0.142
0	1	0	1	1	0	0	11	0.593	0.310
0	0	1	0	0	0	0	9	0.422	0.156
0	1	1	0	1	0	0	9	0.657	0.365
1	1	1	0	0	1	1	9	0.816	0.681
0	0	0	0	1	0	0	8	0.577	0.258
0	1	1	1	1	0	0	8	0.603	0.297
1	0	0	0	0	0	0	8	0.612	0.365

Table A.1: Truth Table for Positive Outcome Analysis

ICSEA	URBAN	SIZE	TEACH	ATTEND	FUND	OUT	n	Incl.	PRI
1	0	0	0	1	1	0	8	0.716	0.464
1	1	0	0	0	1	1	8	0.831	0.693
1	1	0	1	0	1	1	8	0.806	0.629
0	1	0	1	0	1	0	7	0.642	0.308
0	1	1	0	0	1	0	7	0.667	0.327
1	1	1	1	0	1	1	7	0.802	0.604
0	0	0	0	0	1	0	6	0.490	0.210
0	1	1	0	1	1	0	6	0.791	0.561
0	0	1	0	1	0	0	5	0.553	0.204
1	0	1	0	1	0	0	5	0.697	0.397
1	1	1	0	0	0	0	5	0.673	0.389
0	0	1	0	0	1	?	4	0.485	0.206
0	1	0	0	0	0	?	4	0.516	0.201
0	1	1	1	0	1	?	4	0.665	0.301
1	1	0	1	0	0	?	4	0.672	0.386
1	1	1	1	1	0	?	4	0.796	0.626
0	0	0	0	1	1	?	3	0.630	0.308
0	0	1	1	0	1	?	3	0.486	0.202
0	1	0	1	1	1	?	3	0.769	0.536
1	0	1	0	0	1	?	3	0.647	0.400
1	0	1	1	0	0	?	3	0.604	0.330
0	0	1	0	1	1	?	2	0.629	0.272
0	0	1	1	1	0	?	2	0.539	0.197
0	1	0	0	0	1	?	2	0.684	0.361
0	1	0	0	1	0	?	2	0.659	0.387
1	0	0	0	0	1	?	2	0.621	0.383
1	0	0	0	1	0	?	2	0.685	0.401
1	1	0	0	0	0	?	2	0.687	0.421
0	1	0	0	1	1	?	1	0.798	0.585
0	1	1	1	1	1	?	1	0.781	0.508
1	0	1	1	0	1	?	1	0.622	0.346
1	1	0	0	1	0	?	1	0.802	0.660
1	1	0	1	1	0	?	1	0.785	0.624
1	1	1	1	0	0	?	1	0.667	0.357
0	0	1	1	1	1	?	0	0.610	0.252
1	0	1	0	0	0	?	0	0.609	0.344
1	0	1	1	1	0	?	0	0.670	0.357
1	0	1	1	1	1	?	0	0.700	0.398

Table A.2: Truth Table for Negative Outcome Analysis

ICSEA	URBAN	SIZE	TEACH	ATTEND	FUND	OUT	n	Incl.	PRI
1	1	1	0	1	1	0	148	0.328	0.103
0	0	0	1	0	0	1	82	0.812	0.755
0	1	0	1	0	0	1	29	0.914	0.872
0	1	1	1	0	0	1	29	0.910	0.863
0	0	0	1	1	0	0	19	0.798	0.679
1	0	0	1	0	0	0	19	0.718	0.577
1	1	0	0	1	1	0	18	0.460	0.155
0	0	0	1	0	1	1	17	0.828	0.750
1	1	1	1	1	1	0	17	0.476	0.154
0	1	1	0	0	0	1	15	0.896	0.834
1	0	0	1	1	0	0	14	0.719	0.514
0	0	0	1	1	1	1	13	0.807	0.668
1	0	0	1	0	1	0	13	0.746	0.607
1	1	0	1	1	1	0	13	0.516	0.209
0	0	0	0	0	0	1	12	0.864	0.798
1	0	0	1	1	1	0	12	0.714	0.509
1	0	1	0	1	1	0	12	0.774	0.539
1	1	1	0	1	0	0	12	0.625	0.349
0	0	1	1	0	0	1	11	0.900	0.858
0	1	0	1	1	0	1	11	0.816	0.689
0	0	1	0	0	0	1	9	0.893	0.844
0	1	1	0	1	0	1	9	0.801	0.631
1	1	1	0	0	1	0	9	0.608	0.319
0	0	0	0	1	0	1	8	0.853	0.742
0	1	1	1	1	0	1	8	0.831	0.701
1	0	0	0	0	0	0	8	0.777	0.634
1	0	0	0	1	1	0	8	0.755	0.536
1	1	0	0	0	1	0	8	0.618	0.306
1	1	0	1	0	1	0	8	0.668	0.366
0	1	0	1	0	1	1	7	0.840	0.692
0	1	1	0	0	1	1	7	0.837	0.672
1	1	1	1	0	1	0	7	0.698	0.396
0	0	0	0	0	1	1	6	0.863	0.788
0	1	1	0	1	1	0	6	0.732	0.438
0	0	1	0	1	0	1	5	0.885	0.795
1	0	1	0	1	0	0	5	0.800	0.602
1	1	1	0	0	0	0	5	0.792	0.611
0	0	1	0	0	1	?	4	0.866	0.794
0	1	0	0	0	0	?	4	0.878	0.799
0	1	1	1	0	1	?	4	0.855	0.699
1	1	0	1	0	0	?	4	0.793	0.613
1	1	1	1	1	0	?	4	0.659	0.374
0	0	0	0	1	1	?	3	0.836	0.692
0	0	1	1	0	1	?	3	0.870	0.798
0	1	0	1	1	1	?	3	0.733	0.464
1	0	1	0	0	1	?	3	0.765	0.600
1	0	1	1	0	0	?	3	0.805	0.670

Table A.2: Truth Table for Negative Outcome Analysis

ICSEA	URBAN	SIZE	TEACH	ATTEND	FUND	OUT	n	Incl.	PRI
0	0	1	0	1	1	?	2	0.860	0.725
0	0	1	1	1	0	?	2	0.887	0.803
0	1	0	0	0	1	?	2	0.821	0.639
0	1	0	0	1	0	?	2	0.784	0.612
1	0	0	0	0	1	?	2	0.765	0.617
1	0	0	0	1	0	?	2	0.789	0.599
1	1	0	0	0	0	?	2	0.772	0.578
0	1	0	0	1	1	?	1	0.715	0.414
0	1	1	1	1	1	?	1	0.774	0.492
1	0	1	1	0	1	?	1	0.800	0.654
1	1	0	0	1	0	?	1	0.617	0.340
1	1	0	1	1	0	?	1	0.643	0.376
1	1	1	1	0	0	?	1	0.815	0.643
0	0	1	1	1	1	?	0	0.869	0.748
1	0	1	0	0	0	?	0	0.795	0.656
1	0	1	1	1	0	?	0	0.817	0.643
1	0	1	1	1	1	?	0	0.801	0.602

A.2 HSEE Further Tables

Table B.3: List of States with an HSEE and Their Year of First Administration

State	Type of Test	Year(s) first Administered	Year Diplomas First Withheld	Grade First Administered	Grade(s) Exam Aligned to
Alabama	SB	1984, 1995	1985, 2001	10	11
Alaska	SB	2000	2004	10	8 to 10
Arizona	SB	1999	2006	10	10
Arkansas	EOC	2001, 2010	N/A	Algebra 1	Algebra 1
California	SB	2001, 2004	2006	10	10, Algebra 1
Florida	SB	1998	2003	10	10
Georgia	SB	1991	1994	11	9 to 11
Idaho	SB	2004	2006	10	10
Indiana	SB	1997	2000	10	9
Louisiana	SB	2001	2003	10	9 to 12
Maryland	EOC	2001	1989, 2009	DOS	10
Massachusetts	SB	1998	2003	10	10
Minnesota	SB	1996, 2010	2000, 2010	9 to 11	8 to 10
Mississippi	EOC	2000, 2007	2006	DOS	9 to 11
Nevada	SB	2001	2003	10	9 to 12
New Jersey	SB	1991, 2002	2003	11	11
New Mexico	MC	2011	2012	11	9 to 12
New York	RE	1878, 2000	2003	DOS	9 to 12
N. Carolina	EOC	2006	2010	DOS	Course-specific
Ohio	SB	1990, 2005	1994, 2007	10	10
Oklahoma	EOC	2001	2012	DOS	HS standards
Oregon	SB	2009	2012	3	11
Rhode Island	CO	2012	2012	11	9, 10
S. Carolina	SB	1986, 2005	2006	9	9
Tennessee	EOC	2001	2005	DOS	10
Texas	SB	1990, 2003	2005	11	HS standards
Virginia	EOC	1998	2004	DOS	Course-specific
Washington	SB	1999, 2010	2008, 2010	10	10

Notes: Data collected by the authors. SB means standards-based, EOC means end of course, DOS means depends on subject, RE means Regents examination, and N/A means not available.

Table B.1: Effect of HSEE on Graduation Rate (All States), Non-Parametric Specifications

Variables	Graduation Rate			
	Coefficient	Standard Error	Coefficient	Standard Error
	[1]	[2]	[3]	[4]
$(t - t_i^*)$	0.121*	(0.045)		
$(t_i^* - 10)$			-1.710	(1.126)
$(t_i^* - 9)$			-2.020	(1.143)
$(t_i^* - 8)$			-2.045	(1.161)
$(t_i^* - 7)$			-1.793	(1.226)
$(t_i^* - 6)$			-1.433	(1.231)
$(t_i^* - 5)$			-1.638	(1.163)
$(t_i^* - 4)$			-1.834	(1.268)
$(t_i^* - 3)$			-2.551*	(1.237)
$(t_i^* - 2)$			-3.617**	(1.315)
$(t_i^* - 1)$			-3.644**	(1.069)
(t_i^*)	-2.227**	(0.657)	*Base Category*	
$(t_i^* + 1)$	-2.487**	(0.765)	-3.379**	(1.023)
$(t_i^* + 2)$	-2.370**	(0.912)	-3.193**	(1.082)
$(t_i^* + 3)$	-2.324*	(1.073)	-3.076*	(1.197)
$(t_i^* + 4)$	-1.857	(1.197)	-2.487	(1.427)
$(t_i^* + 5)$	-1.433	(1.148)	-1.985	(1.342)
$(t_i^* + 6)$	-0.754	(1.341)	-1.185	(1.520)
$(t_i^* + 7)$	-1.015	(1.460)	-1.325	(1.638)
$(t_i^* + 8)$	-1.611	(1.543)	-1.828	(1.739)
$(t_i^* + 9)$	-1.164	(1.759)	-1.285	(2.040)
$(t_i^* + 10)$	-1.384	(1.543)	-1.405	(1.784)
$(t_i^* + 11)$	-1.200	(1.648)	-1.101	(1.871)
$(t_i^* + 12)$	-0.026	(1.844)	0.195	(2.024)
$(t_i^* + 13)$	0.276	(1.644)	0.547	(1.793)
$(t_i^* + 14)$	1.383	(1.123)	1.747	(1.210)
$(t_i^* + 15)$	2.143	(1.237)	2.606*	(1.282)
Intercept	73.554***	(0.267)	74.734***	(0.521)
State FE	YES		YES	
Adjusted R^2	0.790		0.785	
N	1,200		1,200	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

A.3 KNN Further Tables

Table B.2: Effect of HSEE on Achievement (All States), Non-Parametric Specifications

Variables	NAEP Scores (8th Grade Math)			
	Coefficient	Standard Error	Coefficient	Standard Error
	[1]	[2]	[3]	[4]
$(t - t_i^*)$	0.841***	(0.055)		
$(t_i^* - 10)$			-9.300*	(3.984)
$(t_i^* - 9)$			-7.257*	(3.604)
$(t_i^* - 8)$			-8.502*	(3.678)
$(t_i^* - 7)$			-5.884	(4.476)
$(t_i^* - 6)$			-5.366	(3.908)
$(t_i^* - 5)$			-5.266	(3.658)
$(t_i^* - 4)$			-0.776	(3.103)
$(t_i^* - 3)$			-0.650	(4.989)
$(t_i^* - 2)$			-0.874	(3.729)
$(t_i^* - 1)$			-2.563	(4.482)
(t_i^*)	-0.625	(2.057)	*Base Category*	
$(t_i^* + 1)$	0.735	(1.336)	-0.198	(5.730)
$(t_i^* + 2)$	-0.538	(1.194)	0.321	(3.915)
$(t_i^* + 3)$	3.577**	(1.032)	5.981	(3.512)
$(t_i^* + 4)$	0.348	(1.194)	3.465	(3.403)
$(t_i^* + 5)$	3.325*	(1.097)	5.169	(4.195)
$(t_i^* + 6)$	1.748	(1.334)	5.577	(4.065)
$(t_i^* + 7)$	3.018**	(1.045)	8.463*	(3.381)
$(t_i^* + 8)$	1.233	(1.297)	7.050	(3.838)
$(t_i^* + 9)$	2.565	(1.838)	8.611*	(4.150)
$(t_i^* + 10)$	1.582	(1.304)	7.836	(4.069)
$(t_i^* + 11)$	2.696	(1.814)	10.809**	(3.585)
$(t_i^* + 12)$	0.525	(1.254)	8.402	(4.185)
$(t_i^* + 13)$	1.752	(1.702)	10.004*	(3.822)
$(t_i^* + 14)$	-0.016	(1.097)	7.412	(4.135)
$(t_i^* + 15)$	2.294	(1.457)	11.218**	(3.912)
Intercept	270.055***	(0.368)	275.612***	(1.673)
State FE	YES		YES	
Adjusted R^2	0.914		0.617	
N	457		457	

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors are clustered at the State level and robust to heteroskedasticity and serial correlation.

Table C.3: Descriptive Statistics for 2010

N = 1,843	Mean	St. Dev.	Min	Max
Sector	1.840	0.560	1	3
Type	2.052	0.517	1	3
ICSEA Score	1,031.886	72.575	612	1,247
Full-Time Enrollments	419.727	384.398	19.000	3,101.000
Indigenous Enrollments	1.475	3.564	0	100
LBOTE	22.374	23.720	0	100
Teacher-Student Ratio	0.071	0.019	0.047	0.273
Staff-Student Ratio	0.023	0.013	0.002	0.207
Attendance Rate	92.711	2.923	50	100
Funding Per Student	10,461.060	3,696.817	4,032.602	37,349.140
Previous Score	483.728	50.211	344.500	712.000
Current Score	483.722	51.282	343.500	713.500

Table C.4: Descriptive Statistics for 2011

N = 1,879	Mean	St. Dev.	Min	Max
Sector	1.842	0.556	1	3
Type	2.046	0.518	1	3
ICSEA Score	1,032.662	70.824	830	1,239
Full-Time Enrollments	421.082	384.627	7.000	3,271.000
Indigenous Enrollments	1.553	2.859	0	32
LBOTE	20.973	24.785	0	199
Teacher-Student Ratio	0.072	0.020	0.042	0.273
Staff-Student Ratio	0.023	0.013	0.002	0.102
Attendance Rate	92.751	2.782	50	100
Funding Per Student	11,046.180	4,087.142	4,305.501	53,240.710
Previous Score	483.220	51.368	320.000	713.500
Current Score	482.283	49.917	312.500	725.000

Table C.5: Descriptive Statistics for 2012

N = 1,893	Mean	St. Dev.	Min	Max
Sector	1.844	0.557	1	3
Type	2.044	0.518	1	3
ICSEA Score	1,028.696	69.843	831	1,220
Full-Time Enrollments	426.464	387.676	11.000	3,289.000
Indigenous Enrollments	1.684	3.031	0	33
LBOTE	21.213	24.706	0	100
Teacher-Student Ratio	0.073	0.020	0.043	0.266
Staff-Student Ratio	0.025	0.015	0.004	0.210
Attendance Rate	93.481	2.816	54	100
Funding Per Student	11,534.990	4,468.758	2,012.738	68,152.060
Previous Score	482.014	50.192	312.500	725.000
Current Score	480.294	49.746	347.000	723.500

Table C.6: Summary Statistics for Two-Year Data

N = 1,866	Mean	St. Dev.	Min	Max
Sector	1.840	0.557	1	3
Type	2.050	0.517	1	3
ICSEA Score	1,030.907	72.694	612	1,247
Full-Time Enrollments	419.079	383.182	15.000	3,101.000
Indigenous Enrollments	1.522	3.820	0	100
LBOTE	22.718	24.043	0	100
Teacher-Student Ratio	0.071	0.019	0.047	0.273
Staff-Student Ratio	0.023	0.013	0.002	0.207
Attendance Rate	92.689	3.149	39	100
Funding Per Student	10,336.210	3,869.644	0.000	37,349.140
2010 Score	483.089	51.657	320.000	713.500
2012 Score	480.910	49.479	347.000	723.500

A.4 Curriculum Vitae

Personal Details

Caves, Katherine Marie	26 December 1987

Education

September 12 - July 15	Doctoral student at the University of Zurich, Department of Business Administration Chair of Performance Management
August 10 - May 12	Master of Arts, University of Tulsa, Education specialization
August 06 - May 10	Bachelor of Arts, University of California at Berkeley, Interdisciplinary Studies Field Major